Automated 3D Breast Ultrasound Image Analysis

Tao Tan

Automated 3D Breast Ultrasound Image Analysis

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 12 februari 2014
om 10.30 uur precies

door

Tao Tan

geboren op 7 Maart 1984
te Wuhu (China)

Promotor:	Prof. dr. ir. Nico Karssemeijer
Copromotoren:	Dr. ir. Bram Platel
	Dr. Ritse M. Mann
Manuscriptcommissie:	Prof. dr. ir. Bart M. ter Haar Romeny (TUE)
	Prof. dr. Tom Heskes
	Dr. Mireille Broeders

Advanced School for Computing and Imaging

The research described in this thesis was carried out at the Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Center (The Netherlands) and the Advanced School for Computing and Imaging (ASCI) graduate school. ASCI dissertation series number 297.
This work was funded by EU FP7 project HAMAM and Qview Medical Inc (Los Altos, CA).

Financial support for publication of this thesis was kindly provided by the Faculty of Science, Radboud University Nijmegen.

TABLE OF CONTENTS

Introduction

1

1.1 Breast Cancer Screening

Breast cancer is the most frequently diagnosed cancer in women and it is the leading cause of cancer death in women worldwide[48]. Each year 421,000 new cases are diagnosed in the Europe and 129,000 women die from the disease[29]. While causes remain largely unknown, incidence is still increasing in most countries. Currently, approximately 1 in 8 women develops breast cancer during her lifetime. It is known that the survival rate is associated with the stage of the detected cancer. If breast cancer is detected early, mortality is most likely decreased due to more effective treatment and quality of life of patients decreases less because early detection enables less radical surgery and restrictive use of chemotherapy.

Breast cancer screening aims to detect breast cancer in asymptomatic women as early as possible. Current breast cancer screening service programs are performed with mammography. The benefits of mammography-based breast cancer screening to the reduction in cancer mortality have been shown in studies. In 1985, Tabár et al. showed a 31% reduction in mortality from breast cancer and a 25% reduction from more advanced breast cancers in the group invited to screening[92]. Broeders et al.[11] in their review indicated that the best 'European' estimate of breast cancer mortality reduction is 25-31% for women invited for screening, and 38-48% for women actually screened.

However mammography-based breast cancer screening has its limitations. The sensitivity of mammographic screening is seriously impaired in women with dense breasts. The reason is that dense tissues (fibroglandular and stromal tissues) and breast cancer both show equally bright on mammographic images as they have the same or similar X-ray attenuation properties. For extremely dense breasts, a sensitivity as low as 30% has been reported[63]. Therefore, the sensitivity of detecting breast cancer in dense breasts is significantly lower than that in fatty breasts[79,80]. Moreover, for female carriers of BRCA1/2 mutations, exposure to diagnostic radiation before age 30 is associated with an increased risk of developing breast cancer[78].

Women with dense breasts have a risk of breast cancer four to six times higher than that of women with no or little dense tissue[7]. To improve cancer detection in dense breasts, personalized breast cancer screening with ultrasound (US) has been proposed for women with dense breasts and women with elevated risk factors for developing breast cancer. Supplemental handheld ultrasound (HHUS) breast cancer screening can detect small, early stage invasive cancers that appears to be occult on mammograpms due to breast density[73]. Furthermore it can be used as an alternative screening device to mammography for women with harmful mutations in either BRCA1 or BRCA2 since no radiation is involved. Berg et al.[103] found that the supplemental yield of using ultrasound together with mammography was 4.2 cancers per 1000 women screened with heterogeneous dense tissue in at least 1 quadrant. Corsetti et al.[21] found 37 cancers (0.4%) from 9157 patients with dense breasts (BI-RADS category 3 or 4 density) and negative mammograms. Kaplan et al.[49] found 6 (0.36%) extra cancers from 1862 mammgraphic negative patients with dense breasts.

However handheld ultrasound breast cancer screening is operator dependent, uneasy to reproduce, time consuming and relatively expensive as a screening procedure when performed by radiologists[103]. An alternative to HHUS screening is automated breast ultrasound screening. Automated breast ultrasound scanning is performed by a trained technician or nurse. The transducer is automatically translated across or around the breast to obtain a number of 2D images covering the breast. Volumetric breast images are generated by stacking the 2D images together and the US volumes are stored enabling temporal comparison.

Automated breast ultrasound is a radiation free technique and it is relatively inexpensive and effective. Kelly et al.[53] showed that by adding automated breast ultrasound screening to mammography the sensitivity increased from 40% to 81% in asymptomatic women with heterogeneous and extremely dense breasts. However, after the addition of automated breast ultrasound, recall rate increased from 4.2% to 7.2%. Giuliano et al. performed a clinical trial in two groups of women with mammographically dense breasts[37]. The control group was offered digital mammography and the test group was offered additional automated 3D breast ultrasound to digital mammography. They found a sensitivity of 97.67% in the test group and 76% in control group. Both studies were with one year follow-up.

1.2 Automated Breast Ultrasound Acquisition Systems

There are different commercial automated 3D breast ultrasound systems available. Most common 3D breast ultrasound systems are SonoCiné ultrasound system, Sonix-Embrace automated breast ultrasound, Somo-V automated 3D breast ultrasound system, ACUSON S2000 automated breast volume scanner and SVARATM warm bath ultrasound breast imaging system. These systems mainly produce a series of 2D ultrasound slices by manually or automatically moving a 2D transducer array across or around the breast with or without compression and some systems further generate a 3D volume in which the breast can be visualized in three orthogonal views. The width and frequency of the transducer also differ from system to system resulting in different image resolutions. Moreover, the patient positioning for imaging is also different.

1.2.1 SonoCiné Ultrasound System

The SonoCiné Ultrasound system (Sono-ciné, Inc., NV) was used in the study of Kelly et al.[53]. The images are collected with multi-frequency transducers (7 MHz to 12 MHz). The transducer is attached to a computer-guided mechanical arm, and transversal images are acquired in longitudinal rows, overlapping 7 to 10 mm for complete coverage. The width of the rows without the overlap was about 4.2 to 4.5 cm. In most women the number of rows varied from 4 to 7 for each breast. During imaging, the mechanical arm controls transducer speed and position, with a trained ultrasound technologist maintaining appropriate contact pressure and orientation vertical to the skin. The space between recorded 2D transversal images is 0.8 mm. Typical imaging time is 10 to 20 min for each participant, with additional participant preparation time of 5 to 10 min.

1.2.2 Automated 3D Breast Ultrasound

The automated 3D breast ultrasound system (ABUS) is a novel 3D breast ultrasound imaging device. Imaging involved frontal compression of the breast by a dedicated membrane, allowing a wide transducer with a large width to move smoothly over the breast surface. In the ABUS scanning device the transducer is translated with a constant speed across the breast to obtain a 3D volume of imaging data covering a large segment of the breast. Depending on the size of the breast, up to five views are acquired for each breast. Positioning and compression are to some extent standardized and include anterior-posterior, lateral, medial, superior or inferior views. Two types of ABUS systems are currently available on the market: the Somo-V automated 3D breast ultrasound system developed by U-systems (Sunnyvale, CA, USA) and the ACUSON S2000 ABUS system (Fig. 1.1) developed by Siemens Medical Solutions (Mountain View, CA, USA). The device by U-systems generates an image with a maximal size of 14.6 cm by 16.8 cm on the coronal plane and a maximal depth of 4.86 cm while the device by Siemens generates an image with a maximal size of 15.4 cm by 16.8 cm on the coronal plane and a maximal depth of 6 cm. The two devices differ in two aspects. Firstly, the frequency of the transducer by Siemens is between 5.0 and 14.0 MHz, which is adjustable to the breast size, while the frequency of the transducer by U-systems is fixed at 8.0 MHz or 10.0 MHz. Secondly, the compression of the membrane on the breast is manually applied using the device by U-systems, while the compression is applied automatically by the device by Siemens using a mechanical arm.

In the latest ABUS system by U-systems, a curved transducer is adopted instead of a flat transducer to match the natural curved shape of breasts. It helps to create a uniform compression across the entire breast and enable deep penetration. Fig. 1.2 illustrates this type of curved transducer.

Figure 1.1: The ACUSON S2000 Automated Breast Volume Scanner by Siemens.

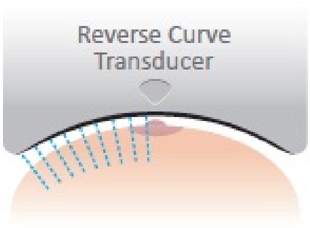


Figure 1.2: An illustration of the reverse curve soft touch transducer by U-systems. The transducer is curved to match breast shapes

1.2.3 SonixEmbrace Automated Breast Ultrasound System

The SonixEmbrace automated breast ultrasound system (Fig. 1.3) developed by Ultrasonix uses a 360 degree, rotating concave ultrasound transducer, to capture detailed B-mode ultrasound images of the uncompressed breast. These images can be reconstructed to view the breast structures in a coronal plane. The patient lies comfortably in a prone position. Each patient is positioned in the same way so the transducer is able to automatically and consistently capture images, enabling reproducible results during follow-up exams. The disadvantage of this system is that it images uncompressed breasts which means more penetration depth is required and lower frequency has to be used to achieve large scanning depth. BSonixEmbrace won the title of Best New Radiology Device by Auntminnie in 2012 (http://www.auntminnie.com/) .

Figure 1.3: The SonixEmbrace automated breast ultrasound system by Ultrasonix.

1.2.4 SVARATM Warm Bath Ultrasound Breast Imaging System

The SVARATM warm bath ultrasound breast imaging system (TechniScan, Salt Lake City, UT) is a software-controlled ultrasound system that performs an automated whole breast examination. The patient lies prone on the examination table with the breast suspended in a warm water bath maintained near skin temperature, while an ultrasound array is moved in a continuous scan so that a complete series of 2D image slices of the breast are collected. Scans take approximately 12 minutes per breast and required no breast compression.

Table 1.1 summarizes the main imaging properties of different 3D ultrasound systems. The automated 3D breast ultrasound (ABUS) by U-system or Siemens has its advantages compared to other 3D ultrasound systems. The translation of the transducer is fully automated, and the imaging protocol is standardized. It use a wide transducer (around 15 cm) to cover large segments of the breast for each image. Moreover, it provides 3D breast volume in which we could inspect the coronal plane of the breast. It uses wide high frequency transducer resulting in high-quality images which

Table 1.1: Characteristics of different automated ultrasound systems

System Type	shown in 2D/3D	fully automated ?	transducer frequency	transducer width	compression	patient position	scanning time per-breast
SonoCiné	2D	no	7 - 12 MHz	5.2 cm	yes	supine	10-20 mins
SonixEmbrace	3D	yes	?	?	no	prone	2-4 mins
Somo-V	3D	no/yes	8.0 MHz/10.0 MHz	14.6 cm	yes	supine	6-10 mins
ACUSON S2000	3D	yes	5.0 MHz -14 MHz	15.4 cm	yes	supine	6-10 mins
SVARATM	3D	yes	?	?	no	prone	12 mins

is possible because of the frontal compression which makes the breast as thin as possible and the scanning time per-breast (6 views to 10 views) is relatively short. In this thesis, ABUS is used to perform studies.

1.3 CADe and CADx systems in Breast Ultrasound

With ABUS systems the effort and costs of acquiring images are lower, compared to the costs of hand held ultrasound imaging, since the imaging procedure can be performed by radiographers, as opposed to hand-held ultrasound which has to be performed by radiologists. Despite this, implementation of ABUS in screening will lead to a strong increase of the workload of radiologists, because reading ABUS screening exams is much more time-consuming than reading mammograms. Apart from that, in the volumetric ABUS images, small lesions may be overlooked more easily than in mammograms. To support search and interpretation, computer aided detection (CADe) systems can be used, in which computers identify potential locations of abnormalities with high sensitivity. The use might also be helpful to speed up the workflow.

In screening mammography, computer-aided detection systems have been considered as a tool to relieve the shortage of trained readers for double reading[4,76] and it can also be used as an aid to help the first reader to localize the abnormality more quickly and more accurately, suppressing the chance that lesions are overlooked by radiologists due to fatigue or the lack of experience. Different studies[9,36,72,86] have shown that use of CADe can improve the reader performance of breast cancer detection or screening. In the USA the majority of mammograms is nowadays read with CADe support. However, a CADe system for ABUS is still absent or still in a infant stage. In this thesis we will describe a CADe system for detecting breast cancers in ABUS systems in which dedicated techniques are introduced to deal with region segmentation in ABUS and specific features differentiating cancers from false positives are adopted.

Detecting lesions in breast ultrasound can be separated into two stages: a first stage of determination of the initial lesion candidate and a second stage of a region classification to reduce false positives. Most existing CADe systems differ most in the first stage. In 3D ultrasound, Moon et al.[69] used Hessian analysis based multi-scale blob detection to initially locate tumors and subsequently combined blobness, internal echo and morphology features to perform region classification. A sensitivity of 70% was achieved with 2.7 false positives per image. Ikedo et al.[46] applied Canny edge detector to detect near-vertical edges and used the edges as a cue for searching the tumors. Chang el al[16] applied image registration to merge three-pass images into full-view images. Subsequently, lesion candidates were determined by using simple intensity-based thresholding in combination with a set of region criteria. In 2D US, Drukker et al.[23,24] used a radial gradient index filter to determine possible tumor regions. Those regions were further segmented using a region-growing technique by maximizing the average radial gradient on the contour and then regions were classified using Bayesian neuron-network classifier based on a number of region features.

A potential negative effect of ABUS screening, however, may be an increase in the number of false positives. Therefore, it is important to develop techniques to help radiologists to minimize the number of false positive recalls. A computer aided diagnosis (CADx) system can help readers to assess the probability that a certain lesion is malignant, which in turn might reduce the amount of recalls and biopsies for benign lesions. Several CADx systems have been developed to automatically classify lesions in 2D breast ultrasound[17,24,44,61,70] and 3D breast ultrasound[70,85]. A CADx system includes three stages: segmentation of breast lesions, feature extraction and classification. Segmentation methods such as active contour, level set, watershed, dynamic programing etc. are applied to deal with partial defined boundary or surface of lesions in ultrasound. From the lesion segmentation effective features can be extracted from the shape, margin, posterior acoustic behavior, internal echo pattern, texture, coronal spiculation or retraction pattern, etc. Those features are combined using supervised learning techniques to generate a likelihood of each lesion being a cancer. It should be noted that CADx deals the classification between benign and malignant lesions when a lesion is detected while the second stage of CADe deals with the classification between malignant lesions and non-lesion regions.

1.3.1 Observer Studies

The aim of developing computer aided diagnosis (CADx) systems is to help radiologists diagnose lesions more effectively when these lesions are already located by a reader or a computer-aided detection (CADe) system. There are many studies showing that by using CADx systems, the reader performance can be improved in breast imaging[64,101,105]. In 3D ultrasound, Sahiner et al.[84] already showed that the use of CAD improved the radiologists' accuracy of distinguishing malignant from benign breast masses on images from an experimental ultrasound system from which the image quality is not comparable to ABUS images. In this thesis, we are going to study whether a developed CADx system would benefits diagnostic performance of readers in ABUS.

1.4 Thesis Outline

This thesis describes methods for quantitative analysis of automated 3D breast ultrasound images. The main focus is automated lesion detection and classification. Other techniques are also developed to serve the main goal including image normalization, lesion segmentation, chestwall segmentation, etc.

The outline of this thesis is as follows. Chapter 2 provides an automated segmentation method for breast lesions and describes the development of a CADx system for the classification of breast lesions in ABUS. In particular, we investigated if the CADx system can be improved by incorporating coronal spiculation features. In chapter 3, the performances of readers with and without the aid of a CADx system is compared. Chapter 4 presents a fully automatic approach to segment the chestwall in ABUS which can be used to suppress false positives in cancer detection and used as landmarks for image registration. Chapter 5 describes a two-stage detection system for breast cancers. Chapter 6 describes an automated method to determine corresponding lesions in target views when a lesion was already pointed out by readers or CADe system in a source view. The thesis will be finalized by a summary and discussion chapter

Computer-aided Lesion Diagnosis in ABUS

2

Tao Tan, Bram Platel, Henkjan Huisman, Clara I. Sánchez, Roel Mus, and Nico Karssemeijer

Original title: Computer-aided Lesion Diagnosis in Automated 3D Breast Ultrasound Using Coronal Spiculation

Published in: IEEE Trans Med Imaging. 2012 May;31(5):1034-42

Abstract

A computer-aided diagnosis (CAD) system for the classification of lesions as malignant or benign in automated 3D breast ultrasound (ABUS) images, is presented. Lesions are automatically segmented when a seed point is provided, using dynamic programming in combination with a spiral scanning technique. A novel aspect of ABUS imaging is the presence of spiculation patterns in coronal planes perpendicular to the transducer. Spiculation patterns are characteristic for malignant lesions. Therefore, we compute spiculation features and combine them with features related to echotexture, echogenicity, shape, posterior acoustic behavior and margins. Classification experiments were performed using a support vector machine (SVM) classifier and evaluation was done with leave-one-patient-out cross-validation. Receiver Operator Characteristic (ROC) analysis was used to determine performance of the system on a dataset of 201 lesions. We found that spiculation was among the most discriminative features. Using all features, the area under the ROC curve (A_z) was 0.93, which was significantly higher than the performance without spiculation features (A_z=0.90, p=0.02). On a subset of 88 cases, classification performance of CAD (A_z=0.90) was comparable to the average performance of 10 readers (A_z=0.87).

2.1 Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in females worldwide[48]. To reduce the mortality of breast cancer screening programs have been introduced. Because early detection reduces the risk of metastatic disease these programs are a key to successful treatment. Currently, mammography is the primary modality for breast cancer screening and diagnosis. As a complementary modality, targeted 2D ultrasound (US) is widely used for diagnosing lesions, while most breast biopsies are performed with ultrasound guidance. For diagnosis, ultrasound has long been recognized as a valuable technique for characterization of cysts[42,47,57].

With new generations of the technology, the quality of US has strongly improved, increasing its potential for screening applications. In particular, there is an increasing interest to use US for screening in women with dense breasts, because of the poor performance of mammography in these women[5,55]. Ultrasound has a high sensitivity for detecting invasive cancer in dense breasts. However, disadvantages of handheld ultrasound in screening are operator-dependency, a long examination time, and the inability to acquire and archive 3D volumetric images of the breast. Automated ultrasound scanning approaches are developed to alleviate these problems. In a recent study Kelly et al.[53] used an automated scanning technique operated by technicians, resulting in 2D image sequences covering the entire breast which are presented to the radiologists as a movie. Detection sensitivity increased significantly when compared with mammography alone in dense breasts. Limitations of 2D ultrasound can be overcome by automated 3D breast ultrasound (ABUS). This new modality provides 3D ultrasound images of large sections of the breast from the skin surface to the chest wall. The modality involves compression of the breast using a dedicated membrane and a wide transducer mounted in a scanning device. Fig. 2.1 shows a 3D breast image generated by ABUS. Typically three or four views per breast are acquired that target different areas of the breast.

Computer-aided diagnosis (CAD) systems have been developed for different breast imaging modalities to support radiologists in detecting and diagnosing breast lesions more accurately and efficiently[14,64,84,86]. In screening mammography CAD is nowadays widely used. Development of CAD in 2D breast ultrasound has received much attention in the literature[2,13,20,35,54,61], but applications have not yet been introduced in clinical practice. It is expected, however, that interest in breast ultrasound CAD will strongly increase when the technique is getting used in screening. Computerized features of breast lesions that have been reported are related to lesion shape, margin, posterior acoustic behavior, intensity distribution (echogenicity and echo texture).

Due to limited availability of 3D acquisition devices, only few investigators have reported work in 3D breast ultrasound. Wenkel et al.[107] showed that the use of ABUS achieved a high sensitivity in a selected patient group. Sahiner et al.[85] developed a CAD system based on extracted texture and morphological features of lesions in 3D US volumetric images and it was shown that the use of this CAD system[84] significantly improved radiologists' accuracy in distinguishing malignant from benign breast masses. The data were obtained with an experimental 3D breast ultrasound device, in which lesions were imaged after identification of their location using 2D ultrasound and mammography and the images were acquired by manual translation of the transducer. Chen et al.[19] extracted the texture characteristics of lesions from 3D US images obtained by an automated scanner using a transducer with a scan width of 40 mm. The 3D US imaging in this study also required identification of lesion location using 2D ultrasound. Ikedo et al.[46] developed an automatic scheme for detection of masses in whole breast ultrasound images in which 3D characteristics were not used due to large interval between slices (2 mm). Chang et al.[16] developed a computer-aided detection system to automatically detect lesions from their whole breast images which were obtained by stitching three series of 2D US images together. Moon et al.[71] applied the level set method to segment lesions and proposed a CAD system based on 3D texture and morphology features using ABUS images, which yield an area under

Receiver Operator Characteristic curve A_z of 0.95 on a dataset of 147 lesions. It should be noted that A_z values from different studies can not be directly compared because results depend strongly on the difficulty of the datasets.

In this paper, we present results of computerized analysis of lesions in ABUS data. A novel aspect of this modality is the ability to image high quality coronal views of the breast, representing slices in parallel to the skin surface. These views are strikingly different from regular B-mode ultrasound scans which radiologists are familiar with. In particular, spiculation patterns surrounding malignant lesions are often observed in coronal views, while spiculation cannot be observed well or is completely absent in 2D handheld ultrasound. This additional value of coronal views of 3D breast ultrasound was previously mentioned in[83] and[106]. In this paper we focus on automated classification of benign and malignant lesions, extending work previously reported[93]. In particular, we investigate features to quantify spiculation in 3D ultrasound and to assess their value in classification of benign and malignant lesions. To our knowledge, this is the first paper in which the use of features extracted from coronal planes in breast ultrasound in a CAD system is investigated.

To assess the performance of the presented CAD system we performed experiments using a set of lesions with known pathology. To compare standalone performance of the system to that of radiologists an observer study was conducted in which a subset of the cases was presented to experienced breast imaging specialists.

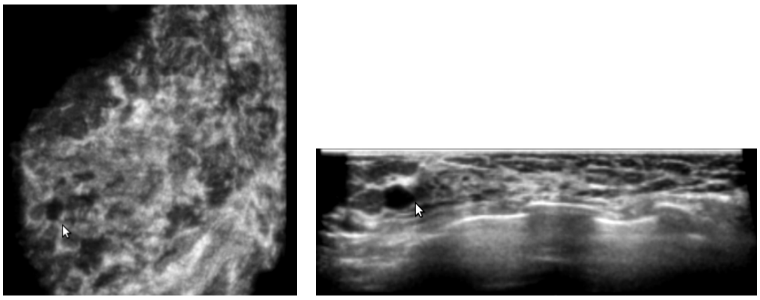

(a) (b)

Figure 2.1: A benign lesion in ABUS (indicated by the arrow). Two slices through the lesion are shown, a coronal slice (a) which is obtained from transforming the original stack of transversal images and a transversal slice (b) which is acquired by the probe at a particular location. The latter view is similar to a regular hand-held ultrasound image.

2.2 Materials and Methods

2.2.1 Dataset

The breast ultrasound images used in this study are a representative sample of diagnostic cases obtained in routine clinical care from Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands), Falun Central Hospital (Falun, Sweden), and the Jules Bordet Institute (Brussels, Belgium). All data was anonymized. Institutional Review Board (IRB) approval was not required. All the centers used the automated breast ultrasound system developed by U-systems, Inc. (SomoVu, Sunnyvale, CA, USA). Imaging involved frontal compression of the breast by a dedicated membrane, allowing a wide transducer with a width of 14.5 cm to move smoothly over the breast surface. The systems used were equiped with 8 Mhz transducers. In the ABUS scanning device the transducer is translated with a constant speed across the breast to obtain a 3D volume of imaging data covering a large segment of the breast. Depending on the size of the breast, up to four views are acquired for each breast. Positioning and compression are to some extent standardized and include anteroposterior, lateral, medial or superior views. Each 3D volumetric view was generated with an in-plane pixel size of 0.285 mm by 0.285 mm or 0.134 mm depending on the settings of the scan and a slice thickness of 0.6 mm. For processing, images were resampled to obtain 0.6 mm cubic voxels.

In this study, a complete dataset of ABUS studies from 158 patients with abnormalities was used. Among 610 views of those 158 patients, 179 views in which lesions are visible were used and annotated. Views in which a lesion is on the boundary of the image domain were not annotated and were excluded. In total, 201 lesions were annotated, including 85 malignant lesions and 116 benign lesions. Table 2.1 shows the distribution of lesions by histology types in the complete dataset and also the subset we used for an observer study (described in subsection 2.2.6). All malignant lesions were confirmed by biopsies and 38 benign lesions were confirmed by biopsies and the remaining benign lesions were confirmed by image interpretation alone or by combining information from different modalities.

2.2.2 Intensity normalization

Images in our dataset were acquired over a longer period of time (2006-2010) and in three institutes. Due to variations of scan parameter settings and use of different software versions the intensity levels of tissue types (fatty tissue, dense tissue, etc) varied from patient to patient. Therefore, before lesion segmentation and feature computation, intensity normalization was performed on all images.

Table 2.1: Number of lesions of different histology types in the complete dataset and observer study (OS) dataset.

Histology	Complete dataset	OS dataset
Malignant lesions		
Infiltrative ductal carcinoma	60	38
Ductal carcinoma in situ	3	1
Infiltrative lobular carcinoma	7	6
Other	15	2
Total	85	47
Benign		
Cyst	47	9
Fibroadenoma	24	14
Fibrocystic change	24	11
Other	21	7
Total	116	41

In ABUS imaging, US signal can only be acquired if a good contact between the probe and the skin exists. Therefore only voxels in columns under the area with good contract hold valid information. The valid area on the skin is shown on coronal view in the resulted image. To be able to normalize intensities inside the breast in one image, a 2D breast mask was generated in the coronal plane. The same 2D mask is suitable for all coronal slices in the volume.

To this end, a coronal projection was obtained by averaging a series of 2D coronal slices. The breast mask was obtained by applying Otsu's thresholding method[77] on the coronal projection. To avoid a negative effect of slices through the skin and thoracic volume on the normalization, only slices with depth between 10% and 50% of the total depth of the image volume were used in the projection.

After thresholding, in the projection, the boundary of the largest foreground connected component (breast) was refined by dilation and erosion operations. Holes inside the boundary caused by lesions or shadows were filled, obtaining the breast mask shown in Fig. 2.2(a).

Tissues in the breast, by rough approximation, can be categorized into fatty and dense tissue. We use this for normalization of the intensities in the 3D volume. In the volume, within the breast mask, again using Otsu's thresholding, voxels with depth between 10% and 50% of the total image depth are categorized into fatty tissue and dense tissue (Fig. 2.2(b)). Subsequently, the median intensities of fatty tissue and dense tissue are extracted: I_f and I_d. The normalized intensity I for each voxel is computed

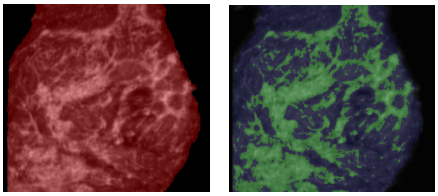

(a) (b)

Figure 2.2: (a) 2D breast mask overlay (red) in a coronal plane. (b) Fatty tissue overlay (blue) and dense tissue overlay (green).

as:

$$I = \frac{(I'_d - I'_f)(I_o - I_d)}{I_d - I_f} + I'_d \tag{2.1}$$

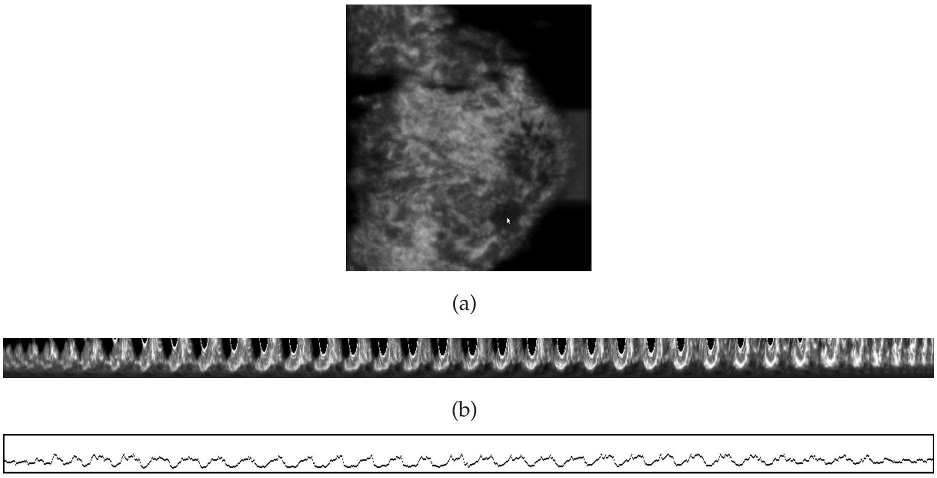
where I_o is the original voxel value and I'_d and I'_f are constants representing reference values for fatty tissue and dense tissue.

2.2.3 Lesion segmentation

For classification of lesions a quantitative representation of lesion features has to be defined. As computation of such features generally requires an accurate delineation of the abnormality, segmentation is an important step in lesion classification algorithms. Various segmentation methods have been proposed to segment lesions in different modalities[1,20,99,104]. In ultrasound, inhomogeneity of intensities inside lesions, background structure, partially undefined boundaries, and different posterior acoustic behavior of lesions, make lesion segmentation a difficult task. To obtain a reliable and computationally efficient lesion segmentation method we extended a spiral-scanning based dynamic programming technique, which was originally introduced by Wang et al.[104] for pulmonary nodule segmentation in CT.

The proposed method can be summarized as follows: First, using the center of the annotated lesion as a seed point, a volume of interest (VOI) is created by defining a sphere with a fixed radius (6 cm) at the lesion center. Second, by applying a spiral scanning technique[104] to the VOI, a 2D resampled image is generated. Third, an optimal path in the resampled 2D image which delineates the lesion boundary is determined by minimizing a cost function by dynamic programming[99,104]. An example is shown in Fig. 2.3. To increase robustness, Wang's original approach is extended by using three spiral scans in orthogonal orientations instead of just one, resulting in three optimal paths. In this way the method is less dependent on the orientation of the spiral scanning. From these three paths edge point candidates are sampled in 3D. The final lesion segmentation is obtained from these candidates by voxel classification as described in[104]. The classifier determines whether voxels are part of the lesion by taking into account the distance of the nearest edge candidate points to the lesion center.

An advantage of the proposed method is that it can be adapted to particular segmentation tasks by modification of the cost function determining the optimal path. While using the intensity gradient to detect edges locally, leakage of the segmentation can be prevented by means of a smoothness term in the cost function[104]. In addition, the cost function can be used to incorporate prior knowledge of the lesions to be segmented. In our cost function we used the expected radius (9 mm) of breast lesions[99]. This is important for lesions with ill defined boundaries and dark posterior shadowing. Fig. 2.4 shows the segmentation result of the same lesion shown in Fig. 2.3. In this example, the segmentation method manages to determine a good approximation of the surface of the lesion, despite the ill-defined boundaries,



(a)

(b)

(c)

Figure 2.3: (a) A 3D ultrasound volume with the cursor indicating a malignant lesion in a coronal slice. (b) shows the resampled 2D image obtained from one spiral scanning. (c) shows the optimal path delineating lesion boundary computed by the dynamic programming algorithm.

2.2.4 Feature computation

Given a segmentation V_l of a lesion, features can be computed to represent lesion properties as echotexture, echogenicity, shape, posterior acoustic behavior and margins. These features are used for classification and should thus represent relevant aspects

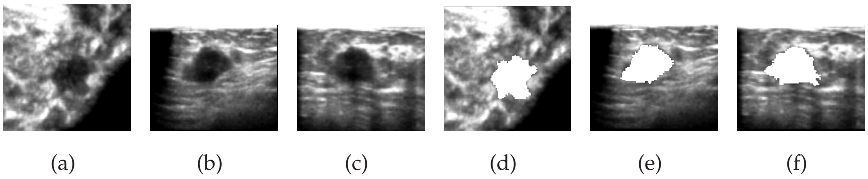


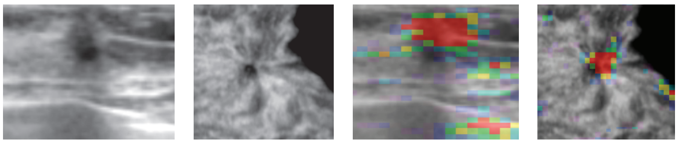
Figure 2.4: Example of a segmentation result. (a) (b) and (c) show the coronal, transversal, and sagittal view of a malignant lesion respectively. (d), (e) and (f) shows the segmentation overlay (white) in coronal, transversal and sagittal views respectively.

of breast lesions in ultrasound as good as possible. To define features, we have used descriptions in the radiological literature as a guide[18,19,33,43,58,82–84,91,106]. Table 2.2 lists the 11 features investigated in this study and their abbreviations.

Table 2.2: lesion features and their abbreviations.

Feature	Abbrevations
the mean of the spiculation values	SP_{mean}
the 90th percentile of the spiculation values	$SP_{90\%}$
the 75th percentile of the spiculation values	$SP_{75\%}$
variance of intensities	VI
entropy	H
average intensity	AI
margin contrast	MC
volumetric height-to-width ratio	$VHWR$
sphericity	SH
compactness	CP
posterior acoustic behavior	PAB

Spiculation patterns are common in mammography in association with invasive breast cancer, but are not observed well or are completely invisible in 2D ultrasound. Interestingly, however, spiculation is often observed on coronal slices of 3D ultrasound images[58,83,106] (Fig. 2.5(b)). To extract features representing spiculation we use a method originally developed for mammography[50,51]. In this method, in each voxel i a measure of linear concentration is computed from a 2D line orientation map in the coronal plane through i. The line orientation map is computed by maximizing the directional second order Gaussian derivative at each voxel j over its orientation in the coronal plane. Indicating the line orientation at voxel j by θ_j, the maximized magnitude of the directional derivative W_j has orientation $\theta_j + \pi/2$. To measure the concentration of line structures pointing to a central region of radius R at voxel i, a circular neighborhood is used as shown in Fig. 2.6. All voxels j with a distance r_{ij} ($r_{ij} \in [r_{min}, r_{max}]$) within the coronal planes to i are selected when the magnitude of W_j exceeds a small threshold. This selected set of voxels is denoted by N_i.



(a) (b) (c) (d)

Figure 2.5: Figures (a) and (b) show a malignant lesion in transversal and coronal views, respectively. Spiculation can be observed well in the coronal view. Figures (c) and (d) show the lesion with the spiculation feature map in color overlay. The spiculation value of the overlay increase with the color changing from blue to red.

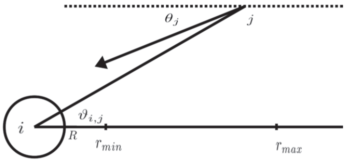


Figure 2.6: Voxels j that are located at a distance $r_{r,j}$ between r_{min} and r_{max} and that are directed toward a disk of radius R centered at the test site i make contributions to spiculation calculation.

For any voxel j in N_i, a contribution x_j of voxel j to the concentration measure at site i is defined as:

$$x_j = \begin{cases} 1 - p_j & \text{if } h(\theta_j, \vartheta_{i,j}, r_{i,j}) = 1 \\ -p_j & \text{else} \end{cases} \tag{2.2}$$

where p_j is the probability that voxel j would be pointing to the site i in a random pattern, defined by

$$p_j = \frac{2R}{\pi r_{i,j}} \tag{2.3}$$

and $h(\theta_j, \vartheta_{i,j}, r_{i,j})$ determines whether voxel j is oriented towards the center i, which is defined as:

$$h(\theta_j, \vartheta_{i,j}, r_{i,j}) = \begin{cases} 1 & \text{if } |\vartheta_{i,j} - \theta_j| < \frac{R}{r_{i,j}} \\ 0 & \text{else} \end{cases} \tag{2.4}$$

where $\vartheta_{i,j}$ is the orientation of the line between voxel j and voxel i and θ_j is the line orientation at voxel j.

Using binomial statistics, the normalized spiculation measure[50] at i is defined by

$$f = \frac{\sum_{j \in N_i} x_j}{\sqrt{\sum_{j \in N_i} p_j (1 - p_j)}} \tag{2.5}$$

To make this measure less dependent on the choice of the neighborhood size (r_{max}), it is computed as a function of r_{max} (12 mm $< r_{max} <$ 30 mm). The parameters r_{min} and R are both set to 3 mm. The maximum f_m of f over the range of neigborhood sizes (r_{max}) was chosen as the spiculation value of voxel i.

The spiculation value f_m is computed for every voxel in the ABUS image, thus yielding a spiculation map in every coronal slice. An example is shown in Fig. 2.5. To extract lesion spiculation features from the voxel spiculation values, we computed the mean (SP_{mean}), the 90th percentile ($SP_{90\%}$) and 75th percentile ($SP_{75\%}$) of the spiculation values of voxels inside the lesion.

Echotexture and echogenicity of a lesion are often good indicators for the malignancy of a lesion[18,19,33]. Variance of intensities (VI), entropy (H) and the average intensity (AI) of voxels inside a lesion are computed to characterize these properties. Entropy (H) is defined as:

$$H = -\sum_i p_i \cdot \log_2(p_i) \tag{2.6}$$

where p_i is the normalized histogram count of bin i computed from the intensities of voxels inside the segmented lesion.

There is usually a sharp demarcation between benign lesions and the surrounding tissue while malignant lesions often have indistinct margins[82]. We calculate a feature termed margin contrast (MC) by taking the difference of the average intensities in the inner and outer border of the lesion. The inner border consists of voxels inside the lesion whose distance to the lesion boundary is no more than than 1.2 mm. The outer border consists of voxels outside the lesion within 1.2 mm of the boundary.

In breast ultrasound, it is well known that lesions that are wider than that they are tall, are more likely to be benign[82,91]. This property is related to higher compressibility of benign lesions. To describe the wider-than-tall shape of lesions in ABUS, we introduce volumetric height-to-width ratio ($VHWR$) which is calculated as:

$$VHWR = \frac{h}{w} \tag{2.7}$$

where h, the height, is defined by the difference of the maximal depth of the lesion z_{max} and the minimal depth z_{min}. The width w is defined as the effective diameter of the coronal cross section of the lesion through the center, i.e. the diameter of the circle that has the same area as the cross section.

Due to different growth rates of breast lesions, malignant lesions tend to have an irregular shape and benign lesions, on the other hand, tend to have a spherical or oval shape[82]. Sphericity (SH) and compactness (CP) are used to characterize the irregularity of lesion shapes. Sphericity is defined as:

$$SH = \frac{|V_l \cap S|}{|S|} \tag{2.8}$$

where S is a sphere at the lesion center which has equal volume as the segmented lesion. Compactness is defined as:

$$CP = \frac{A^3}{|V_l|^2} \tag{2.9}$$

where A is the area of the surface of V_l.

The posterior acoustic behavior of a lesion is important for the classification as sound waves are attenuated differently in softer and harder tissues. Malignant lesions are harder than benign lesions on average[28,56]. Therefore, in the posterior regions of lesions, benign solids tend to have no shadow, cysts usually have enhancement, and malignant lesions tend to have a dark shadow. The posterior acoustic behavior is a commonly analyzed feature in breast ultrasound[43,85]. In this study, the posterior acoustic behavior feature (PAB) is computed as the difference between the average intensities of the region posterior to the lesion and the region surrounding the posterior region with the same depth range. The feature is defined as:

$$PAB = \bar{I}_p - \bar{I}_s \tag{2.10}$$

where $\bar{I}_p$ and $\bar{I}_s$ are the average intensities of the posterior region V_p and the surrounding region V_s respectively, which are defined as:

$$\begin{aligned} V_p = \{(x, y, z)|(d(x, y) < 0) \\ \wedge(z_{max} < z < z_{max} + l)\} \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} V_s = \{(x, y, z)|(0 < d(x, y) <= 2) \\ \wedge(z_{max} < z < z_{max} + l)\} \end{aligned} \tag{2.12}$$

where l is the pre-defined height of the posterior region, which is taken equal to the height h of the lesion in this study, z_{max} is the maximal depth of the lesion, and $d(x, y)$ is the signed distance map transformed from the binary 2D projection in coronal view of the lesion. In the signed distance map, the absolute value of a pixel represents the shortest distance of that pixel to the region boundary. Pixels inside the border boundary are assigned negative values and pixels outside the boundary are assigned positive values.

2.2.5 Training and evaluation

To automatically generate a score that indicates the malignancy of a lesion, a support vector machine (SVM) with a radial basis function (RBF) as the kernel was used as the classifier. In order to evaluate the classifier performance and avoid possible bias, a nested leave-one-patient-out cross validation scheme was adopted. The inner cross-validation loop is used to perform the tuning of the SVM and kernel parameters using a grid search process while the outer loop is used to compute the estimate of the system performance. To investigate the benefits of incorporating coronal spiculation features, the experiments were performed using all features combined and using all features except spiculation features. The discriminative performance of the classification was evaluated by computing the area under Receiver Operating Characteristics (ROC) curve denoted as A_z. All ROC curves shown in this paper were fitted using the proper binormal model[66]. Performances of the CAD system with and without using spiculation features were compared using DBM MRMC 2.3[22,41]. Statistical analysis was performed using the fixed-reader with random-cases model. To investigate the effectiveness of each individual feature, we computed A_z values for each feature separately. To investigate how well the CAD system performs with solid masses using all features, we also performed ROC analysis on the 154 solid masses in our dataset, excluding the 47 cysts.

2.2.6 Observer study

To determine how well the CAD system performs compared to human readers, an observer study was conducted. From the complete dataset, a total of 88 patients were included in this study, where for each patient one ABUS view with a lesion was selected. Cases for the study were manually selected to form a dataset with a roughly balanced number of malignant and benign lesions. During the selection, there is no preference on certain types of lesions except that most cases with cysts were left out because these are very easy to interpret. In the study set, 47 lesions were malignant and 41 were benign. The distribution of lesion types is shown in Table 2.1. Ten readers (6 radiologists and 4 residents) were invited to read the series of cases. Three radiologists and two resident had extensive experience with ABUS image interpretation. The remaining three radiologists and two residents had limited or no experience with ABUS but had extensive experience with interpreting 2D ultrasound images. Readers rated the lesions on a continuous scale ranging from 0 (benign) to 100 (malignant). Before the study sessions, readers were presented a series of 26 training cases with 23 benign and 14 malignant lesions which were also randomly selected from the complete dataset of 201 lesions, to familiarize them with the 3D viewing workstation. Performance of the readers was compared with that of standalone CAD using DBM MRMC 2.3[22,41], treating both readers and cases as random samples.

2.3 Results

Table 2.3 shows A_z values and the corresponding standard deviation obtained by the bootstrap method[27] using individual features on all 201 lesions. The A_z value of individual features varied from 0.51 to 0.84. Margin contrast (MC) and volumetric height-to-width ratio ($VHWR$) were the most discriminative features, both with an A_z value of 0.84. Three spiculation features (SP_{mean}, $SP_{75\%}$ and $SP_{90\%}$) led to A_z values of 0.83, 0.83 and 0.81. Among features characterizing the irregularity of lesion shapes, compactness (CP) had the best performance (A_z=0.82). Posterior acoustic behavior feature (PAB) had an Az value 0.80. Scatter plots in feature space of MC and SP_{mean} and in feature space of MC and $VHWR$ are shown in Fig. 2.7 and Fig. 2.8 respectively.

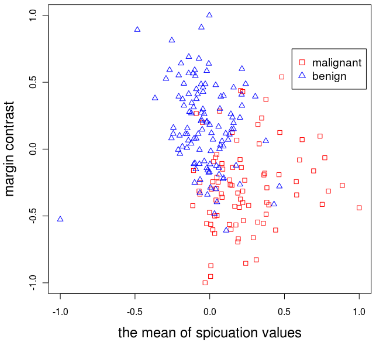


Figure 2.7: Scatter plot in feature space of the mean of spicualtion values (SP_{mean}) and margin contrast (MC).

The ROC curves of the CAD system with and without spiculation features on the dataset of all lesions are shown in Fig. 2.9. The CAD system achieved an A_z value of 0.93 using all the features and an A_z value of 0.90 without using spiculation features. The performance increase with spiculation was statistically significant (p = 0.02). On the complete dataset, it can be seen that the sensitivity of the CAD system using all features retained even high when the false positive (FP) fraction is reduced from 1 to 0.4. At a FP rate of 0.4, the sensitivity is 0.98. On the dataset of solid lesions, the A_z value of the CAD system using all features was 0.90±0.03.

Results of the reader performances from the observer study are shown in Table 2.4. The ROC curves of the CAD system and all radiologists are shown in Fig. 2.10. On the

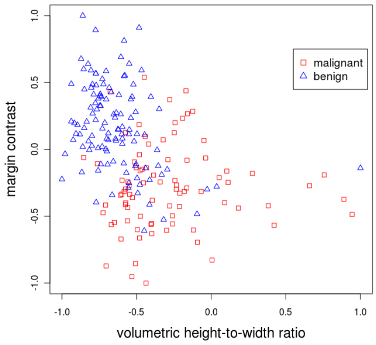


Figure 2.8: Scatter plot in feature space of volumetric height-to-width ratio ($VHWR$) and margin contrast (MC).

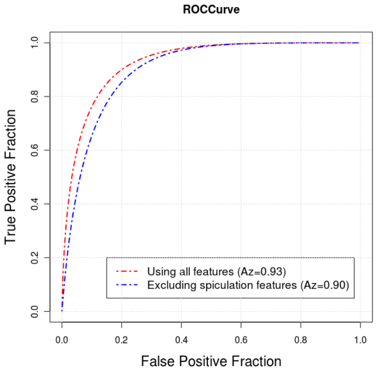


Figure 2.9: Performances of the CAD system with and without spiculation features.

Table 2.3: Performance of feature(s).

Feature(s)	A_z	Std Dev
MC	0.84	0.03
$VHWR$	0.84	0.03
SP_{mean}	0.83	0.03
$SP_{75\%}$	0.83	0.03
CP	0.82	0.03
$SP_{90\%}$	0.81	0.03
PAB	0.80	0.03
H	0.72	0.04
VI	0.61	0.04
AI	0.59	0.04
SH	0.51	0.04
$All\ features$	0.93	0.02
$All\ without\ spiculation\ features$	0.90	0.02

selected study cases our CAD system achieved an A_z value of 0.90±0.04. The mean A_z values of the radiologists, the residents, and all readers were 0.88±0.03, 0.85±0.04 and 0.87±0.03 respectively. The standard deviation for the mean Az values of the readers was obtained by treating readers and cases as random samples[22,41]. The difference between the performance of CAD and all readers was not significant (p=0.26). Neither was the difference between the performance of CAD and the residents (p=0.14).

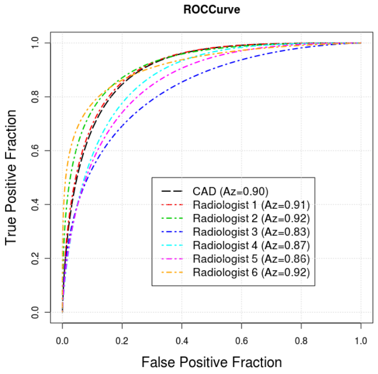


Figure 2.10: Performances of the CAD system and the radiologisits.

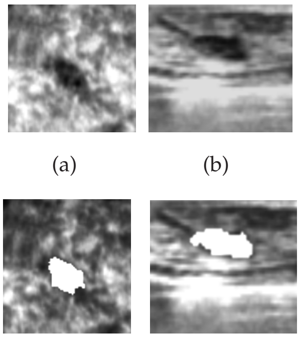
Table 2.4: Performance of the radiologists (RAD) and the residents (RES) in terms of A_z values.

Reader	A_z	Std Dev
RAD1	0.91	0.03
RAD2	0.92	0.03
RAD3	0.83	0.04
RAD4	0.87	0.04
RAD5	0.86	0.04
RAD6	0.92	0.03
Mean RAD	0.88	0.03
RES1	0.87	0.04
RES2	0.77	0.05
RES3	0.86	0.04
RES4	0.89	0.03
Mean RES	0.85	0.04
Mean of All	0.87	0.03

2.4 Conclusion and Discussion

A computer aided diagnosis system was developed for the classification of malignant and benign lesions in automated 3D breast ultrasound (ABUS). In the system, following automated lesion segmentation 11 features were computed to capture relevant lesion characteristics. Using a support vector machine as the classifier in combination with leave-one-patient-out cross-validation, an A_z value of 0.93 was obtained on a dataset of 201 lesions. The CAD system retained a sensitivity of 0.98 when the false positive rate was reduced from 1 to 0.4. This indicates that the CAD system has potential to improve diagnosis and that it also might be used to be more selective with biopsies in diagnostic procedures. It was found that the performance of the CAD system significantly improved when spiculation features were used (p=0.02). By conducting an observer study it was found that the CAD system had a performance comparable to that of experienced breast radiologists.

Our results highlight the diagnostic benefits that can be gained by using ABUS compared with regular 2D ultrasound for breast lesion classification. Different to the work by Moon et al.[71], we focused on the contribution of spiculation features derived from ABUS. Spiculation features extracted from coronal planes of ABUS images turn out to be particularly effective, while this feature is difficult to observe in 2D handheld ultrasound. The scatter plot (Fig. 2.7) shows malignant lesions tend to have higher spiculation values. In addition, volumetric height-to-width ratio ($VHWR$) was found



(a) (b)

(c) (d)

Figure 2.11: An example of a malignant lesion which has a oval shape and has no shadowing in its posterior region. (a) and (b) show the lesion in coronal and transversal view respectively. (c) and (d) show the white overlay of the lesion segmentation in coronal and transversal view respectively.

to be one of the most discriminating features with an A_z value of 0.84. The effectiveness of this feature may result from the modest compression of the breast in ABUS. Differences in compressibility of benign and malignant lesions can result in lesion shape differences.

For some lesions the CAD system did not perform well. Fig. 2.11 shows a malignant lesion to which the CAD system assigned a low probability of being malignant. In the shown coronal plane, the segmentation failed to cover some of the lobes extended from the mass. Reasons might be that the intensity level of those lobes was different from that of the central region of the lesion. Also the smoothness constraint in the cost function may have prevented the lobes to be included in this case. With its oval shape and the absence of spiculation and posterior shadowing the CAD system rated this lesion as most likely benign, even though the inhomogeneity of intensity distribution inside the lesion was suggestive for malignancy. In our study, we found that variance of intensities (VI) and entropy (H) characterizing echotexture were not effective as single features. Reasons might be that those features are computed without considering spatial information of voxels inside the lesion. The performance of the CAD system might be improved by incorporating more advanced texture features.

Lesion segmentation plays a crucial role in the CAD system since the computation of features related to lesion shape is largely dependent on the accuracy of segmentations. Due to ill-defined borders of breast lesions in ultrasound, in particular in the posterior regions, segmentation is often difficult. In this study we used a dynamic programming approach, because this is known to be an effective approach in situations where large sections of lesion boundaries are missing. In future work, it would be worthwhile to experiment with other segmentation approaches and to compare these to the method we used here.

Our study shows that margin contrast (MC) is an effective feature in the classification task. This confirms the usefulness of computerized features extracted from lesion margins in breast ultrasound[2,35,43]. The computation of this feature depends on the choice of the width of inner border and outer border. Thus far we did not study the effect of border width parameters. Computing margin contrast at different scales (border widths) might be a feasible way to improve performance.

The Az value of the CAD system is higher than the average Az value of the observers. It should be noted, however, that most observers were lacking experience in reading ABUS images. Although they were familiar with the transversal sections displayed in the study, since these are similar to regular breast ultrasound, most readers did have little experience with interpreting lesions in coronal planes. In the study we failed to show a significant difference between the performance of CAD and the observers. This may be due to the fact that the power of the observer study was relatively low. We did use a fair number of lesions, but many were relatively easy to discriminate and therefore did not contribute much to the statistical power. Further, the CAD system is designed to assist radiologists. It is necessary to investigate whether the performance of observers improves with using CAD.

The CAD system was developed and validated with ABUS images from one manufacturer, while the technology is still under development. Newer systems use transducers with a higher frequency, resulting in a higher resolution. However, as long as some compression is used, the characteristics of lesions remain similar in 3D breast ultrasound imaging. For example, benign lesions are more likely to have a wider-than-tall shape. Furthermore, the presence of spiculation patterens or architectural distortions associated with breast cancer were reported in 3D breast ultrasound images acquired with different ultrasound systems[58,83,106]. Moreover, the intensity normalization we used makes feature computation less dependent on acquisition settings. Therefore, we expect that the system we developed is applicable to newer automated breast ultrasound systems, although retraining may be needed with data from these systems.

Acknowledgment

This work has been funded by the EU funded project HAMAM (IST-2007-224538) within the Seventh Framework Programme (FP7). The authors would like to thank André Grivegnée from Cancer Prevention and Screening Clinic, Jules Bordet Institute, Brussels, Belgium and László Tabár from Department of Mammography, Falun Central Hospital, Sweden for providing the data.

Evaluation of the Effect of Computer-Aided Classification on Reader Performance

3

Tao Tan, Bram Platel, Thorsten Twellmann, Guido van Schie, Roel Mus, André Grivegnée, Ritse M. Mann, and Nico Karssemeijer

Original title: Evaluation of the Effect of Computer-Aided Classification of Benign and Malignant Lesions on Reader Performance in Automated 3D Breast Ultrasound

Published in: Academic Radiology 2013;20(11):1381-1388

Abstract

Rationale and Objectives are to investigate the effect of a newly developed computer-aided diagnosis (CAD) system on reader interpretation of breast lesions in automated 3D breast ultrasound. A CAD system was developed including automated lesion segmentation in 3D, extraction of lesion features such as spiculation, margin contrast, and posterior acoustic behavior, and a classification stage. 88 patients with breast lesions were included for an observer study. 47 lesions were malignant and 41 were benign. Eleven readers (7 radiologists and 4 residents) read the cases with and without CAD. We compared the performance of readers with and without CAD using receiver operating characteristic (ROC) analysis. The CAD system had an area under ROC curve (AUC) of 0.92 for discriminating benign and malignant lesions, where the unaided reader AUC ranged from 0.77 to 0.92. Mean performance of inexperienced readers improved when CAD was used (AUC = 0.85 vs 0.90; p = 0.007), while mean performance of experienced readers did not change with CAD (AUC=0.89). By using CAD system for classification of lesions in automated 3D breast ultrasound, which on its own performed as good as the best readers, performance of inexperienced readers improved while that of experienced readers remained unaffected.

3.1 Introduction

Mammography is the primary modality for breast cancer diagnosis and screening. However, the sensitivity for breast cancer in dense breasts is lower than fatty breasts[80]. To improve cancer detection in dense breasts, ultrasound can be used as an adjunct modality[53 6]. Schaefer et al. showed that compared to using mammography alone, by combining mammography and ultrasound, 15.9% extra breast cancers were revealed in dense breasts[89]. However, conducting an exam with handheld ultrasound is cumbersome, user dependent and time-consuming and it has limitations of visualizing breasts in 3D. These disadvantages were alleviated with the introduction of automated 3D breast ultrasound (ABUS). ABUS provides a 3D view of the breast and allows reconstruction of coronal planes in which spiculation patterns associated with invasive breast cancers are often observed[83],[106]. These characteristic patterns cannot be observed well or are completely absent in 2D ultrasound. Image acquisition of ABUS can be performed by non-radiologists in a standardized protocol, allowing batch reading by radiologists, as is common practice in screening. A potentially negative effect, however, may be an increase in the number of false positive findings, either due to benign lesions or due to artifacts. Therefore, it is important to develop techniques to help radiologists to minimize the number of unnecessary recalls, by aiding them to distinguish between benign and malignant lesions..

Computer systems can aid radiologists both in the detection and diagnosis. A computer aided detection system has the potential to help readers to quickly locate potential lesions and reduce the amount of missed lesions. A computer aided diagnosis (CAD) system can help readers to assess the probability that a certain lesion is malignant, which in turn might reduce the amount of recalls and biopsies for benign lesions. A complete system incorporates both functions, they are however developed independently. This paper focuses on the classification task .

CAD systems have been developed extensively for mammography and 2D US[98 44 17 25]. For ABUS, Moon et al.[71] proposed a CAD system based on 3D texture and morphology features. Previously we reported on the development of a CAD system in ABUS in which spiculation analysis is incorporated[94]. The system was demonstrated to have a standalone performance similar to that of radiologists. In this work we investigate in a reader study if such a system can be used to aid radiologists with lesion classification.

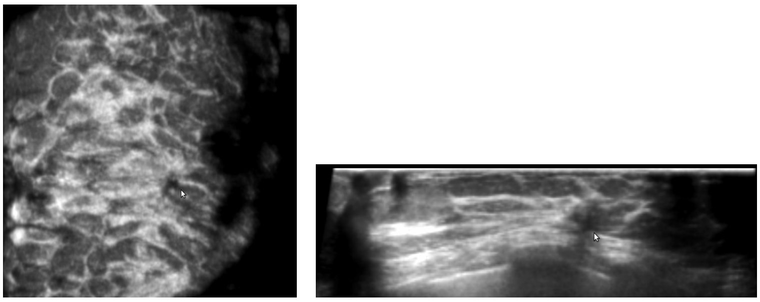
The purpose of this work is to compare the performance of readers in the task of classifying breast lesions as malignant or benign with and without the aid of CAD.

3.2 Materials and Methods

3.2.1 Dataset

The institutional review board approved this retrospective study and waived informed consent. The breast ultrasound images used in this study are a representative sample of diagnostic cases obtained in routine clinical care from three different medical centers. All the centers used the SomoVu automated 3D breast ultrasound system developed by U-systems, Inc. (Sunnyvale, CA, USA). For all acquisitions a linear 8 MHz/10 MHz transducer with a width of 14.5 cm was used. During imaging, patients were in supine position and the breast was slightly compressed by a dedicated membrane. The scanning device automatically translates the transducer in the cross-plane direction during acquisition. Each 3D volume was generated with an in-plane resolution of 0.285 mm by 0.285 mm or 0.134 mm depending on the software version and machine settings and a slice thickness of 0.6 mm. Figure 3.1 shows a transversal view and a coronal reconstruction of a breast with a malignant lesion .

By the time we conducted this observer study, 201 lesions (with a radius 9.6 +/- 3.8 mm) in 158 patients who underwent ABUS imaging between 2006 and 2010 were available (table 3.1). All malignant lesions were confirmed by biopsies and benign lesions were either confirmed by biopsy or by combining information from different imaging modalities.



(a) (b)

Figure 3.1: A malignant lesion in ABUS (indicated by the arrow). Two slices through the lesion are shown, a coronal slice (a) and a transversal slice (b). The latter is similar to the traditional 2D ultrasound view.

3.2.2 Computer-Aided Diagnosis System

The CAD system used in this study was similar to a previously developed system for lesion assessment described in detail in[94]. However, a different set of features is used for lesion representation, which led to better performance. A brief explanation of the system is given below.

For processing, we resampled the images to obtain 0.6 mm cubic voxels . The input of our computer-aided diagnosis system is a seed point inside a lesion to be diagnosed. The CAD system first segments the lesion using the seed point. Because of inherent image speckles, unclear lesion boundaries, different posterior acoustic behavior, and inhomogeneous lesions, lesion segmentation is a difficult task. We used a dynamic programming approach based on a spiral scanning technique which was initially introduced by Wang et al.[104] and was later adapted by[94] for lesion segmentation in ABUS. The segmentation involves a cost function that deals with edge detection locally[99], prevents leakage of the segmentation by including a smoothness term in the cost function and takes the prior knowledge about the expected mean radius of lesion into account[99].

Given the segmentation of a lesion, features are computed to represent its shape, margin, intensities, and texture. In our previous work[94], we computed features representing spiculation, variance of intensities, echogenicity, and entropy inside the lesion. Furthermore, we computed sphericity, compactness, volumetric height-to-width ratio, margin contrast, and a feature representing posterior acoustic behavior. Among these features, margin contrast and spiculation appeared to be the best features for the classification. In this work we use a different set of features for spiculation and lesion shape as described below.

Table 3.1: Number of lesions of different histology types in the complete dataset and observer study (OS) dataset.

Histology	Complete dataset	OS dataset
Malignant lesions		
Infiltrative ductal carcinoma	60	38
Ductal carcinoma in situ	3	1
Infiltrative lobular carcinoma	7	6
Other	15	2
Total	85	47
Benign		
Cyst	47	9
Fibroadenoma	24	14
Fibrocystic change	24	11
Other	21	7
Total	116	41

Spiculation features are quantitative measures representing converging linear patterns. We compute these by estimating an orientation in the coronal plane for every voxel and by subsequently counting the number of voxels orientated towards a common central region. Features are normalized with respect to a random orientation pattern[51]. An example is shown in figure 3.2. Because computation of the spiculation measure is quite demanding it is not computed at every voxel but at a rectangular grid of selected voxels, resulting in a resolution which is lower than that of the original image. Colored voxels mark locations with a high spiculation, with red indicating the highest values. The spiculation measure has a high value at the center of two-dimensional spiculation patterns. In ABUS this results in a signal in a column in the center of a lesion, since spiculation usually occurs in coronal planes throughout the whole lesion. Therefore, to extract spiculation features we define a cylinder CR with an axis through the lesion center along the depth direction. For the radius of CR we choose a value of 3 mm, which was experimentally determined. The length of the cylinder was determined by the depth of the lesion. We compute the average spiculation value inside the whole cylinder and the maximum average spiculation among coronal slices of the complete cylinder CR. We also computed the average spiculation value inside the upper part of the cylinder above the center of the lesion (see figure 3.3) since we observed that spiculation patterns are generally more visible in the upper part of lesions.

Shape features were included to measure the irregularity of lesions. To compute

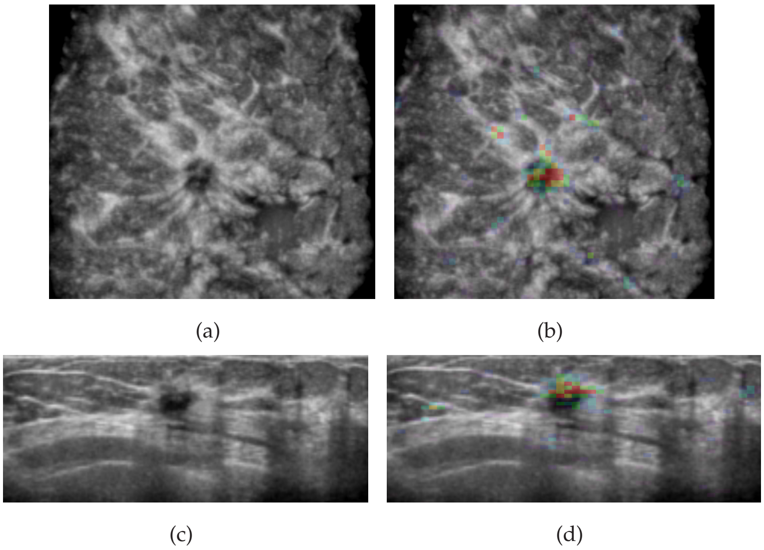


Figure 3.2: (a) and (c) show a malignant lesion in coronal and transversal views, respectively. Spiculation can be observed well in the coronal view. Figures (b) and (d) show the lesion with the spiculation feature map in color overlay. The spiculation value of the overlay increases with the color changing from blue to red.

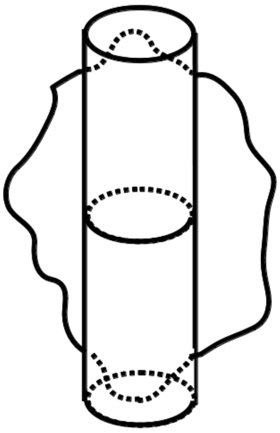


Figure 3.3: In transversal view, a cylinder was created through the center of the lesion along the depth direction and it is evenly divided into two small segments.

these, we register a sphere with the same volume as the segmented lesion to the lesion with an affine transformation. To compare the similarity of the transformed sphere to the obtained segmentation, we computed the Dice coefficient, a well-known similarity measure, representing the size of the union of the segmentation and the sphere divided by the average size of the segmentation and the sphere. We also calculated the volume difference (DF) between the sphere and the segmented lesion. In addition, we included the variance of intensities of the inner border of a lesion (voxels within 1.2 mm of the boundary) and the volume of a lesion as additional features. In total 14 features are used in the CAD system (Table 3.2).

Table 3.2: List of features used in CAD

Features
Average spiculation value of the lesion
Volumetric height-to-width ratio
Margin contrast
Posterior acoustic behavior
Entropy of intensities
DICE coefficient
Echogenicity
Average spiculation value in a cylinder
Average spiculation value of a upper segment in a cylinder
Maximum average spiculation value among coronal slices in a cylinder
Intensity variance inside the lesion
Volume difference between lesion segmentation and the registered result (DF)
Intensity variance of inner border of the lesion
Lesion volume

Linear discriminant analysis (LDA) was used to combine the extracted features. No feature selection step is involved. To train and test the system we performed 10-fold cross-validation on the complete dataset (table 3.1). In this way, for each lesion from the dataset, an unbiased indicator or score of the malignancy was obtained in the range between 0 (benign) and 1 (malignant). In the rest of the paper these are referred to as the CAD scores. As the CAD scores are unbiased due to the cross-validation, these CAD scores can be used in a reader study with the same data to study if lesion classification performance increases with CAD.

3.2.3 Reader Study Design

Eleven readers, (seven radiologists specialized in breast imaging and four residents) were invited for an observer study in which we aimed to compare the diagnostic performance of readers with and without the aid of our CAD system. We also compared the diagnostic performance of readers without CAD to the performance of the CAD system itself (standalone CAD). It is noted that a similar comparison was also performed with the previously described CAD system using a subset of the same reader study data [11]. Three radiologists and two residents had extensive experience with ABUS image interpretation. The remaining four radiologists and two residents had no or limited experience with ABUS but had extensive experience with interpreting 2D breast ultrasound images.

Cases for the observer study were manually selected by a researcher without knowledge of CAD results. The resulting dataset was roughly balanced regarding the number of malignant and benign lesions. The main purpose of the manual case selection was exclusion of obvious cysts, because these are so easy to interpret that they would not contribute to the power of the study. Apart from the cysts, there was no bias in the case selection towards lesion type, lesion size, or breast density. Most ABUS cases in our dataset were obtained from diagnostic procedures in which only one ABUS view was obtained. Therefore, for each patient, only one ABUS view with a lesion was selected. For the cases where multiple views were available, we selected the view where the lesion was located in the most central position. In total 88 patients were included in the study with 47 malignant lesions and 41 benign lesions (table 3.1). Another subset from the complete dataset, containing 26 patients with 23 benign and 14 malignant lesions, was used to train the readers before the experiment. The choice of the subsets was independent from the CAD results. The CAD scores of lesions used in the observer study are obtained with the cross-validation procedure explained in the previous section.

Images were displayed in their original resolution on a dedicated multi-modal breast workstation (HAMAM prototype, MeVis Medical Solutions, Bremen) on a color monitor with a display resolution of 2560 by 1200 pixels. The workstation allowed scrolling through slices with synchronized viewing of coronal, transversal, and sagittal slices. The synchronization could be disabled or enabled by the push of a button. The window level could be automatically or manually adjusted. The lesion to be assessed was pre-identified by a yellow rectangle marker with a fixed size located at the lesion center. The marker could easily be switched on and off to avoid obscuring the lesion while reading. The workstation could be configured so that it would or would not show the associated CAD score next to the marker.

Before the experiment, each reader started with the 26 training patients. The training set served to acquaint the observers with the system, including synchronized orthogonal views, interpreting CAD scores and the controls for adjusting window and level width of the displayed images. During the training program, CAD scores were shown and the ground truth of all the marked lesions was supplied in a printed document for direct feedback after interpretation. After finalizing the training program, the readers continued with the actual experiment.

To suppress the effect of the order of reading modes (with and without CAD) on our results. The observer dataset was randomly split into two parts: set A (45 cases) and set B (43 cases) and he experiment consisted of two separate reading sessions. For the first session, each readers were asked to read set A with CAD scores shown and a second set (set B) without CAD scores. For each patient, a single lesion was shown and readers were asked to rate this lesion on a scale between 0 (benign) and 100 (malignant). After a break (few hours to several days), the experiment continued with the second session. In this session readers were asked to rate the same lesions, but in the second session they read set A without CAD and set B with CAD. Readers were asked to rate these lesions again. Readers were reminded of the importance of using the same scale to rate the lesions during the whole experiment and they were informed that approximately half of all the lesions that were shown in this experiment were malignant and the other half were benign. There was no limitation on reading time.

3.2.4 Data Analysis

Receiver operating characteristic (ROC) was used to analyze the diagnostic performance without and with the aid of CAD for individual readers, as well as to compare reader performance to the standalone performance of the CAD system. The area under the ROC curve (AUC) was used to evaluate the performance. The AUC was obtained with the binormal model[67] using DBM MRMC 2.3, developed by researchers from the University of Iowa and the University of Chicago[22 41]. The significance of differences in reader performance between reading with and without using CAD was analyzed with the Dorfman-Berbaum-Metz method[22 41], which performs an analysis of variance allowing to treat both readers and cases as random samples. Because in our analysis we made multiple comparisons we applied a Bonferroni correction to account for application of multiple statistical tests. A p-value of less than 0.05/N was considered to indicate a statistically significant difference, with N the number of comparisons.

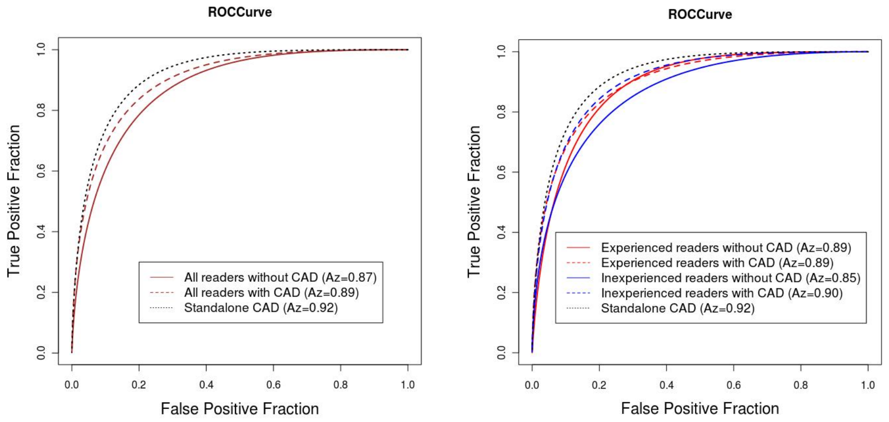
Table 3.3: AUC Values of Performances of Readers without and with CAD

Reader	Without CAD	with CAD
Experienced Reader (p = 0.55)		
Radiologist 1	0.91	0.90
Radiologist 4	0.87	0.86
Radiologist 6	0.92	0.91
Resident 1	0.87	0.92
Resident 3	0.86	0.90
Average	0.89	0.89
Inexperienced Reader (p = 0.007)		
Radiologist 2	0.92	0.93
Radiologist 3	0.83	0.86
Radiologist 5	0.86	0.88
Radiologist 7	0.86	0.90
Resident 2	0.77	0.81
Resident 4	0.89	0.95
Average	0.85	0.90
Reader average	0.87	0.89

3.3 Results

As shown in table 3.3, the AUC of reader performances without CAD range from 0.77 to 0.92 while with CAD the AUC values range from 0.81 to 0.95. By using CAD, the average AUC value remains at a level of 0.89 for experienced readers (with extensive experience on ABUS), increases from 0.85 to 0.90 for inexperienced readers (with no or limited experience on ABUS), and from 0.87 to 0.89 for all readers. Three comparisons were performed, therefore a p-value of less than 0.017 was considered to indicate a statistical significance. The difference in performance with and without CAD was not significant when all readers were pooled, although a trend towards improved performance was observed (p=0.042). The difference was significant for inexperienced readers (p=0.007). Figure 3.4(a) shows the ROC curves of the performance of the readers with and without CAD and standalone CAD. The ROC curves of experienced readers and inexperienced readers are shown in figure 3.4(b).

The AUC value of the performance of standalone CAD was 0.92, which was as good as the best readers.



(a) (b)

Figure 3.4: (a) ROC curves of all Readers without and with CAD and the ROC curve of CAD standalone. (b) ROC curves of the experienced and inexperienced readers without and with CAD.

3.4 Discussion

The performance of the standalone CAD system for ABUS we developed turned out to be excellent on our study material, and compared favorably to that of the readers. It is noted that CAD results are unbiased, because cross validation was used to compute them and because no feature selection was applied. We believe that CAD performs so well because it can exploit information in 3D in an optimal way, while it is trained on a large database. It should be recognized, however, that ABUS is a new modality and that readers are expected to become better with ABUS interpretation when they gain more experience.

Sahiner et al.[84] already showed that the use of CAD improved the radiologists' accuracy of distinguishing malignant from benign breast masses in 3D ultrasound images. In their study an experimental non-automated system was used. Our study confirms that there is a potential benefit of using CAD, though in our study a significant improvement was only obtained for inexperienced readers. With CAD they increased their performance to the level of the experienced readers. This might indicate that the benefits of a CAD system are greater for inexperienced readers than for experienced readers. Results may also be explained by lack of confidence of the experienced readers in CAD. This might change when they have more time to evaluate CAD. For practical reasons, the learning phase in our experiments was short.

From the ROC curves, it can be seen that, with the use of CAD, the average sensitivity of all readers can remain at a high level of 0.98, while the false positive (FP) fraction is reduced from 1 to 0.56. This indicates that use of CAD has potential to improve diagnosis and that for patient care it has potential to be used to reduce false positive recalls in screening or to make better overall decisions with regard to selecting cases for biopsies.

The readers in this study had different levels of experience when it comes to using CAD in their clinical practice. None of the readers had experience with our CAD system. To use it as a classification aid, readers have to combine the CAD score with their own assessment of the lesion. Although readers were trained with CAD before starting the study, it may well be that they were still on a learning curve while performing the experiment.

In the observer study dataset, the number of benign and malignant lesions was balanced. This is different to the real diagnostic situation in which the incidence of benign lesions exceeds that of malignant lesions. However, the effect of this distribution of benign and malignant lesions was the same for reading without and with CAD. Therefore we still could attribute the performance improvement to the use of the CAD system.

This study has some limitations. First, due to practical limitations, in our study we could not accommodate long intervals between reading sessions. In this way, there was an increased risk that readers did memorize cases. We minimized a potential influence of a memory effect by balancing the reading sessions in such a way that each reader evaluated the first half of the cases without CAD and the other half with CAD the first reading session and then the first half with CAD and the other half without CAD in the second reading session. In this way a potential bias of the results towards one of the reading modes was avoided, though there may have been some influence of a memory effect on the difference between the reading modes. This would, however, not affect the conclusions of this study. Second, the dataset used in this observer study consisted of a fair amount of cases that were easy to interpret. As a consequence, the power of the study was lower than we expected. It is also noted that this observer study was conducted using ABUS images obtained with a system from only one manufacturer. ABUS technology is still under development and more recently released systems have higher image quality. To demonstrate the benefit of CAD using newer ABUS systems additional studies will need to be conducted.

In conclusion, without extensive training use of the CAD system we developed only improves the performance of less experience readers in distinguishing malignant from benign lesions in automated 3D breast ultrasound.

Chestwall Segmentation

4

Abstract

In this paper, we present an automatic method to segment the chest wall in automated 3D breast ultrasound images. Determining the location of the chest wall in automated 3D breast ultrasound images is necessary in computer-aided detection systems to remove automatically detected cancer candidates beyond the chest wall and it can be of great help for inter- and intra-modal image registration. We show that the visible part of the chest wall in an automated 3D breast ultrasound image can be accurately modeled by a cylinder. We fit the surface of our cylinder model to a set of automatically detected rib-surface points. The detection of the rib-surface points is done by a classifier using features representing local image intensity patterns and presence of rib shadows. Due to attenuation of the ultrasound signal, a clear shadow is visible behind the ribs. Evaluation of our segmentation method is done by computing the distance of manually annotated rib points to the surface of the automatically detected chest wall. We examined the performance on images obtained with the two most common 3D breast ultrasound devices in the market. In a dataset of 142 images, the average mean distance of the annotated points to the segmented chest wall was 5.59±3.08 mm.

4.1 Introduction

Breast cancer is the most common cancer that affects women, with 421,000 new cases diagnosed in the Europe each year and 129,000 women dying from the disease[30]. While causes remain largely unknown, incidence is still increasing in most countries. If breast cancer is detected early, mortality is decreased due to more effective treatment. Furthermore, the quality of life of these women is maintained because early detection enables less radical treatment. Screening programs have therefore been introduced to detect early breast cancers in asymptomatic women.

The sensitivity of mammographic screening is seriously impaired in women with dense breasts. This because the fibroglandular and stromal tissues (dense tissues) have the same X-ray attenuation properties as tumors and thus both show equally bright on mammographic images. For extremely dense breasts, a sensitivity as low as 30% has been reported in the US[63].

As a complementary modality to mammography, breast ultrasound has a sensitivity which can surpass that of mammography for patients with dense breast tissue[5]. In a recent study, Kelly et al.[53] showed that detection sensitivity increased significantly using 2D image sequences covering the entire breast when compared with mammography alone in dense breasts. Regular two dimensional breast ultrasound, however, suffers from operator dependence, a limited capability to visualize the entire breast and qualified radiologists have to perform this time-consuming task themselves.

These limitations have been technically resolved with the introduction of automated 3D breast ultrasound volume scanning systems which we referred to as ABUS in this paper. The system involves frontal compression of the breast by a dedicated membrane, allowing a wide transducer to automatically move over the breast surface. In the scanning device the transducer is translated with a constant speed across the breast to obtain a 3D volume of imaging data covering a large segment of the breast. Depending on the size of the breast, up to as many as five views are acquired for each breast. Positioning and compression are to some extent standardized and include anteroposterior, lateral, medial, superior, or inferior views. One main advantage of the 3D breast ultrasound system is to view reconstructed images in the coronal plane which offer a complete new way of looking at breast ultrasound and reveal signs of abnormality that could not be observed before, such as very specific spiculation patterns frequently occurring in the presence of cancer[83,106].

However, the reading of these 3D breast ultrasound images is time consuming and demanding for radiologists. Computer-aided detection (CADe) is expected to play an important role in the future to facilitate the reading of these images. CADe suffers from finding many false positives outside of the breast tissue. To remove false positives of a CADe system beyond the chest wall (thoracic volume), determining the location of the chest wall in ABUS images is necessary. Fig. 4.1 shows the separation between thoracic volume and breast tissue by chestwall.

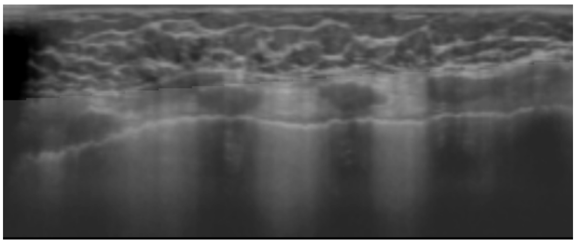
Furthermore, the accurate localization of the chest wall can be used as a landmark for determining spatial correlation between multi-modal images[97]. The spatial correlation is expected to be of great help for fusing information and correlating CADe marks in multi-modal images.

To our knowledge, there is only one paper previously reporting on chest wall segmentation[45], where a deformable volume model is used to determine the chest wall location in ABUS images. The method was only partly validated by determining how well presence of a chest wall in a ABUS scan could be detected, without determining accuracy of the segmentation itself. The performance of the method was limited.

In this paper we describe a novel method to accurately segment the chest wall.

4.2 Dataset

The automated 3D breast ultrasound images used in this study are a representative sample of diagnostic cases obtained in routine clinical care. We used two datasets obtained with two different automated 3D breast ultrasound devices. Dataset A includes 80 patients that were randomly selected from a series of cases obtained using



(a)

Figure 4.1: Separation between thoracic volume and breast tissue by chestwall in transversal plane.

the Somo-V automated 3D breast ultrasound system developed by U-systems (Sunnyvale, CA, USA). Dataset B also includes 80 patients that were randomly selected from cases obtained with the ABUS system developed by Siemens (Erlangen, Germany).

The device by U-systems generates an image with a maximal size of 14.6 cm by 16.8 cm on the coronal plane and a maximal depth of 4.86 cm while the device by Siemens generates an image with a maximal size of 15.4 cm by 16.8 cm on the coronal plane and a maximal depth of 6 cm.

The two devices differ in two aspects. Firstly, the frequency of the transducer by Siemens is between 5.0 and 14.0 MHz, which is adjustable to the breast size, while the frequency of the transducer by U-systems is fixed at 8.0 MHz or 10.0 MHz. Secondly, the compression of the membrane on the breast is manually applied using the device by U-systems, while the compression is applied automatically by the device by Siemens using a mechanical arm. For processing, both types of images were downsampled to obtain 0.6 mm cubic voxels after applying Gaussian smoothing on the original high-resolution image with a scale of 0.3 mm.

For each patient, one single volumetric view was randomly selected. The chest wall was completely invisible in 5 of 80 views in dataset A and in 13 of 80 views in dataset B. Those views in which the chest wall was completely invisible were excluded from this study resulting in 75 views and 67 views in dataset A and dataset B, respectively. The absence of the chest wall in an image may be caused by insufficient scanning depth set by the technician, extremely poor image quality or implants obstructing the chest wall in the image.

4.3 Method

The chest wall is the surface enclosing the ribs and touching the outer rib surfaces. These characteristic rib surfaces are used by human readers to locate the chest wall.

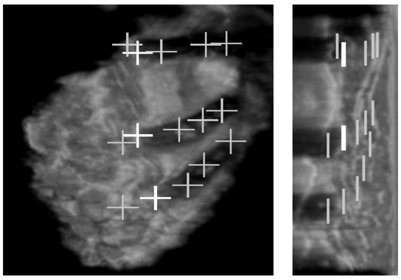

(a) (b)

Figure 4.2: Manually annotated points on the surface of the ribs on coronal view(a) and sagittal view(b). The crosses represent the manually annotated points on the rib surface, projected on the current slice. Annotated points on the current slice are represented by white crosses, and projected crosses are blue.

Therefore, we propose to locate the chest wall by automatically detecting rib surfaces and subsequently fitting a surface through them.

We use multiple features and classifiers to automatically detect points on the surface of the ribs. Subsequently a cylinder is fitted to the resulting point distribution. This cylinder models the local shape of the chest wall in the view at hand.

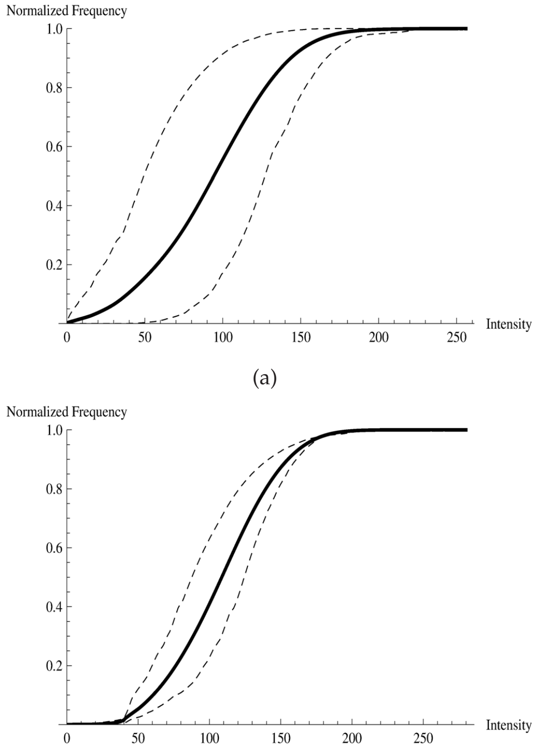
To test our method, we manually marked 10 to 30 rib surface points in the each view of dataset A and B. These points are at the transition between breast tissue and rib shadows for each scan (Fig. 4.2). The annotation was verified with a radiologist.

This section is split into three subsections: the first subsection explains an intensity normalization step applied before chest wall segmentation; the second subsection shows that the partial chest wall visible in a single volumetric view can be modeled by the surface of a cylinder and the third subsection explains a method to automatically detect the rib surfaces.

4.3.1 Intensity normalization

Images in this study were obtained from three different institutes using devices by two different manufactures. Due to variations of scan settings, the intensity levels of the same tissue type (fatty tissue, dense tissue, etc) varied from patient to patient. Therefore, before we perform chest wall segmentation, intensity normalization was performed on all images.

In ABUS imaging, the ultrasound signal can only be acquired if a good contact between the probe and the skin exists. Therefore only voxels in columns under the area with good contact hold valid information. The valid area on the skin is shown on coronal view in the resulting image. To be able to normalize intensities inside the



(a)

(b)

Figure 4.3: The normalized cumulative histogram of intensities within the breast of all the images before normalization (a) and after normalization (b). The Solid line stands for the mean of the normalized cumulative histograms of all images in our dataset and the dashed lines indicate the 95% confidence interval.

breast in one image, a 2D breast mask was generated in the coronal plane. The same 2D mask is suitable for all coronal slices in the volume.

To this end, a coronal projection was obtained by averaging a series of 2D coronal slices. The breast mask was obtained by applying Otsu's thresholding method[77] on the coronal projection. To avoid a negative effect of slices through the skin and thoracic volume on the the mask extraction, only slices with a depth between 10% and 40% of the total depth of the image volume were used in the projection.

After thresholding, in the coronal projection, the boundary of the largest foreground connected component (breast) was refined by dilation and erosion operations. Holes inside the boundary caused by lesions or shadows were filled, obtaining the breast mask.

We use intensity normalization to scale the intensities within the breast mask across all images to a certain range. The standard-score normalization scales all intensities to have a zero mean and a unit stand deviation. However, the imaged breast tissue strongly depends on the composition of the breast, i.e. the proportion of fatty and dense tissue. Therefore standard-score normalization is not appropriate. We previously[94] introduced a normalization method in which the median intensity value of fatty tissue and dense tissue are used as reference values. However this method depends on the accuracy of the tissue segmentation and the intensities of dark shadows are not used. In our current method, within the breast mask, the 98th percentile of all intensity values, I'_{98} and the 2nd percentile of all intensity values, I'_2 were extracted. The normalized intensity I for each voxel is computed as:

$$I = \frac{(I_{98} - I_2)(I_o - I'_2)}{I'_{98} - I'_2} + I_2 \tag{4.1}$$

where I_o is the original voxel value and I_{98} and I_2 are constants representing reference values for the 98th percentile and the 2nd percentile of all intensity values within the breast mask.

We chose the 98th percentile and the 2nd percentile of all intensity values as reference intensities for normalization instead of using the maximum and the minimum of all intensity values to avoid the influence of outliers on the normalization.

Fig. 4.3 shows the mean and the 95 % confidence interval of all normalized cumulative histograms of intensities within the breast before intensity normalization and after. As we can see, before normalization, the distribution of the normalized cumulative histograms spans a much wider range compared to after intensity normalization.

4.3.2 The chest wall as a cylinder

Our method is based on the assumption that the partial chest wall visible in a single volumetric view can be modeled by the surface of a cylinder. To investigate if this conjecture is accurate, we fit the cylinder model to our manually annotated point sets for each scan.

We used a non-linear optimization method[81] to fit a cylinder to the annotated points in 3D space and evaluate the validity of our assumption by observing the average distance of the annotated points to the fitted cylinder surface for each scan.

In all equations in this paper, x and y axis are two orthogonal axis on coronal plane and the z axis (depth direction) is the ventral-dorsal axis. A cylinder in 3D space is fully defined by its radius R and its axis (center line). The axis in turn is defined by a point (x_c, y_c, z_c) and an orientation vector (v_x, v_y, v_z). We set $z_c = 0$ and $v_z = 1$ to remove redundancy. The normalized direction vector with length equal to 1 is denoted by (u_x, u_y, u_z). Thus the parameter vector $\mathbf{p} = \{R, x_c, y_c, v_x, v_y\}$ fully defines the cylinder.

For any point (x, y, z) in 3D space, the closest point on the axis of the cylinder has coordinates (x_a, y_a, z_a) defined by

$$\begin{aligned} x_a &= x_c + D * u_x \\ y_a &= y_c + D * u_y \\ z_a &= z_c + D * u_z \end{aligned} \tag{4.2}$$

where $D = (x - x_c) * u_x + (y - y_c) * u_y + (z - z_c) * u_z$ which is the distance between the point (x_a, y_a, z_a) and (x_c, y_c, z_c) with $z_c = 0$. Therefore, given a set of manually marked rib-surface voxels (x_n, y_n, z_n) $(n = 1, 2, ..N)$ the averaged squared distance to the cylinder surface $ASD(\mathbf{p})$ can be computed by

$$\frac{1}{N}\sum_{n=1}^{n=N}(\sqrt{(x_n - x_a)^2 + (y_n - y_a)^2 + (z_n - z_a)^2} - R)^2 \tag{4.3}$$

where N is the number of annotated points.

We constrain the fitted cylinder by introducing a penalty term $P(\mathbf{p})$

$$P(\mathbf{p}) = w_1 * (\frac{v_x^2 + \xi}{v_y^2 + \xi}) + w_2 * (R - R_{mean})^2 \tag{4.4}$$

in which the first term of $P(\mathbf{p})$ ensures that the orientation of the cylinder remains close to the body axis (y is directed towards to head). ξ is chosen to be a small positive number and is used to ensure a non-zero denominator. The second term penalizes the deviation of the radius of the fitted cylinder from a predefined average radius R_{mean} related to local curvature of the chest wall. The parameters w_1 and w_2 define the weights for each term [1].

The total cost function $C(\mathbf{p})$ is defined by

$$C(\mathbf{p}) = ASD(\mathbf{p}) + P(\mathbf{p}) \tag{4.5}$$

We used Powell's optimization method[81] to find the minimum of the cost function $C(\mathbf{p})$.

The cylinder fitting method was applied to the 75 images of dataset A and the 67 images of dataset B. The average of the mean distance of the annotated points in each image to the fitted cylinder surface for dataset A and dataset B was 2.19 mm±0.99 mm and 2.43 mm±2.76 mm, respectively. Given this small error we conclude that the surface of the chest wall in an ABUS scan can be approximated by the surface of a cylinder.

4.3.3 Chest Wall Segmentation

Our proposed method consists of two stages: first, rib-surface points are identified and second, the previously described fitting procedure is used to fit a cylinder to these points. In 3D breast ultrasound, rib-surface points are at the transition between breast tissue and rib shadows. Detecting rib shadows as landmarks is easier than detection of rib-surface points. Therefore, the extraction of rib-surface points involves two steps: the detection of rib shadows and subsequently the detection of the rib surface.

Detection of Rib Shadows

Due to strong attenuation of the ultrasound signal by the ribs, a shadow is cast behind them in the ABUS images (Fig. 4.2). The detection of these dark sheets aids us in localizing the rib-surface points. Here we describe a set of features we used in combination with a classifier to identify the rib shadows.

Shadow Contrast: we observed that on 2D coronal plane, rib shadows appear as embedded dark horizontal bands. To characterize local contrast for voxel i on coronal plane, we defined a rectangle at center i with a fixed length l (5 mm) and a width h as shown in Fig. 4.4. A shadow contrast feature is defined by:

$$SC(h) = \bar{I}(h + 2d) - \bar{I}(h) \tag{4.6}$$

where $\bar{I}(h)$ represents the average intensity of a rectangle at center i with a length of l and a width of h and d is chosen to be 3 mm. SC is computed as a function of h (5 mm $< h <$ 16 mm) and the maximum of SC is chosen as the feature value. This feature is computed in 2D in the coronal planes. Fig. 4.5 shows this feature as overlay.

---

[1] In our work, we used $R_{mean} = 78mm$, which was an experimental estimate from our dataset. w_1 and w_2 were experimentally set to 100 and 150, respectively.

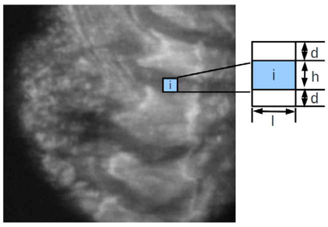


Figure 4.4: For each voxel, a local rectangle is defined to extract local contrast on 2D coronal plane.

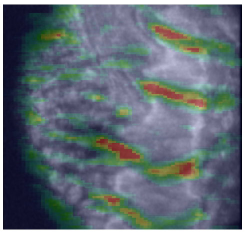


Figure 4.5: A 2D coronal plane with rib shadow contrast (SC) overlay.

Dark Shadow Enhancement: in 3D space, a rib shadow appears as a dark sheet and therefore we apply a sheet detector to enhance the rib shadows in ABUS images. The sheet detector utilizes the eigenvalues and eigenvectors of the Hessian matrix.

We compute the elements of the Hessian matrix by convolving image $I(\mathbf{x})$ with Gaussian kernel

$$L(\mathbf{x},\sigma) = I(\mathbf{x}) \otimes G(\mathbf{x},\sigma) \tag{4.7}$$

where $G(x,\sigma)$ is the Gaussian filter and $L(x,\sigma)$ is the result of the convolution. The normalized Hessian matrix is defined as

$$H(\mathbf{x}) = \sigma^2 \begin{bmatrix} \frac{\partial L(\mathbf{x},\sigma)}{\partial x \partial x} & \frac{\partial L(\mathbf{x},\sigma)}{\partial x \partial y} & \frac{\partial L(\mathbf{x},\sigma)}{\partial x \partial z} \\ \frac{\partial L(\mathbf{x},\sigma)}{\partial y \partial x} & \frac{\partial L(\mathbf{x},\sigma)}{\partial y \partial y} & \frac{\partial L(\mathbf{x},\sigma)}{\partial y \partial z} \\ \frac{\partial L(\mathbf{x},\sigma)}{\partial z \partial x} & \frac{\partial L(\mathbf{x},\sigma)}{\partial z \partial y} & \frac{\partial L(\mathbf{x},\sigma)}{\partial z \partial z} \end{bmatrix} \tag{4.8}$$

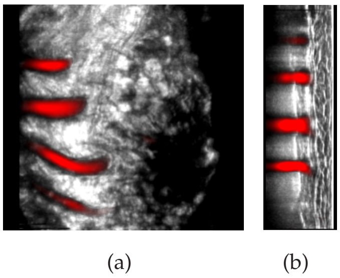
where σ is the scale [1]. The eigenvalues of $H(\mathbf{x})$ are λ_1,λ_2 and λ_3 which correspond eigenvectors V_1, V_2, V_3 where $|\lambda_3| > |\lambda_2| > |\lambda_1|$. The dark sheet enhancement filter[88] is defined as:

$$DS(\lambda_1,\lambda_2,\lambda_3) = \begin{cases} \lambda_3 * e^{-(\lambda_1^2+\lambda_2^2)} & \text{if } \lambda_3 > 0 \\ 0 & \text{if } \lambda_3 < 0 \end{cases} \tag{4.9}$$

The first eigenvector V_3 gives the direction of the greatest curvature. To make the operator more sensitive to sheets perpendicular to the body axis or vertical direction,

---

[1] σ is set to 5 mm which is estimated from the rib-shadow width



(a) (b)

Figure 4.6: A scan with dark shadow enhancement overlay (red). A coronal slice (a) and a sagittal slice (b) are shown.

a weight factor F is calculated as the inner product between V_3 and body-axis vector $(0, 1, 0)$.

Then, the directional dark shadow enhancement image (Fig. 4.6) is computed as

$$SH(\mathbf{x}) = DS(\mathbf{x}) * F \tag{4.10}$$

Additional Shadow Features: the shadow contrast and shadow enhancement alone, however, are not specific enough. Therefore we used a number of additional features to improve the detection of rib shadows including :

- I, image intensity
- $|Gz|$, the absolute values of the first order Gaussian derivative of the intensity image in the depth direction at a scale of 0.6 mm (one voxel) to characterize the fine tissue structure and 2.4 mm (four voxels) to characterize to coarse structures but at a scale small enough to have a low response within the rib shadows
- z, the depth of the voxel

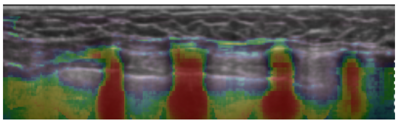


Figure 4.7: A sagittal view image overlaid with the rib shadow likelihood LS. The probability ranges from blue (low) to red (high).

To compute the likelihood that a voxel belongs to a rib shadow, we apply a two-class soft classifier using the features described above as input and using the manual annotations for training.

For the class of rib shadows we used the column of voxels below the annotated rib surface points. For the other class, we used a set of regularly spaced samples (2.4 mm in each direction) that are at least 6 mm above the chest wall (as defined by a fitted cylinder on the manually annotated rib-surface points).

We used an ensemble of five neural networks to classify this data. Each network consists of an input layer of n nodes representing the features, where n is the number of features, and a hidden layer of 8 hidden nodes and an output layer with one node. All networks were initiated with random weights to reduce the influence of the initial weights on the final output of the trained network. The back-propagation algorithm was used for training with a learning rate of 0.005. The final classifier output was computed by averaging the five network outputs. This output, $LS(\mathbf{x})$, is considered to represent the likelihood that a voxel is part of a rib shadow (Fig. 4.7). This likelihood is used as input to the rib-surface detector described below.


(a) (b) (c)

Figure 4.8: (a) A 2D projection obtained by averaging all 2D coronal slices of probability map. (b) A binary 2D projection obtained by thresholding the image (a). (c) Computed contextual feature map.

Detection of Rib Surface

Based on the likelihood map for rib shadows, the extracted dark sheet enhancement feature, and the original intensity image, a number of features are extracted to characterize rib-surface points.

Neighboring Rib Shadows: a 2D coronal projection (Fig. 4.8(a)) is obtained by averaging all 2D coronal slices of the likelihood map. A 2D binary projection (Fig. 4.8(b)) is generated by thresholding the 2D coronal projection. The threshold was defined as the 85th percentile of all likelihood values within the breast mask in each image, as this roughly corresponds to the relative area of the projected rib shadows, ±15 %.

To be able to utilize the contextual information of neighboring rib shadows, for each voxel, a template is defined as Fig. 4.9. The template consists of three rib-shadow segments $S0$, $S1$, and $S2$ with length l (3 mm) and width a (5 mm $< a <$ 15 mm) and two non-shadow segments NS with length l and width b (5 mm $< b <$ 27 mm). Using this template located at each voxel i, the width a and b are locally estimated by:

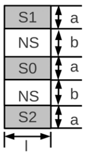


Figure 4.9: A template for extracting contextual information. Blue segments stand for rib shadow regions with a fixed length l and a variable width a and white segments stand for non rib shadow regions in between with a fixed length l and a variable width b. This template is used to compute contextual features indicating the existence of neighboring rib shadows at each pixel in the 2D binary projection.

$$\arg\max_{a,b} \bar{I}_{S0_a \cup S1_a \cup S2_a} - \bar{I}_{NS_b} \tag{4.11}$$

where $\bar{I}_{S0_a \cup S1_a \cup S2_a}$ is the average value of three rib-shadow segments $S0$, $S1$, and $S2$ on the 2D binary projection and $\bar{I}_{NS_b}$ is the average value of non-shadow area NS.

Using the estimated rib shadow width a and non-shadow width b, a contextual feature (Fig. 4.8(c)) for voxel i indicating the neighboring rib shadows (CNR) is defined:

$$CNR = \bar{I}_{S1_a \cup S2_a} - \bar{I}_{NS_b} \tag{4.12}$$

where $\bar{I}_{S1_a \cup S2_a}$ is the average value of the two neighboring shadow segments $S1$ and $S2$ of the segment $S0$ on the 2D binary projection. Since this feature is computed from the 2D binary coronal projection, it has the same value in the z-direction, for each (x, y) location.

Coronal Distance to the Nipple: the distance of voxel (x, y, z) to the nipple in coronal view d_{npl} is extracted. This maybe helpful to reduce false positives, since strong shadowing is caused by the nipple and ducts connecting the nipple. On coronal slices close to the transducer, the nipple is visible as a dark circle. Therefore we applied a Hough circle transform[26] on the coronal slices closer than 4.2 mm to the transducer. The detection results for these slices were fused to obtain the position of the nipple by voting.

Additional Rib-Surface Features: additional features are incorporated characterizing rib-surface. These features are based on the intensity image, the previously described rib shadow likelihood image $LS(\mathbf{x})$ and dark sheet enhancement feature (DS):

- I, image intensity (rib surface appears as bright bands)
- z, the depth of the voxel: it is incorporated since the depth of the rib surface can not be too close to the skin or too deep

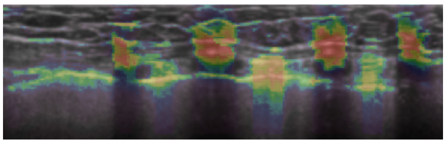


Figure 4.10: A sagittal view image overlaid with the rib-surface likelihood LI. The probability ranges from blue (low) to red (high).

- $Gz_{DS}(x, y, z)$, the maximal value of the first order Gaussian derivative in the depth direction on the original feature image $DS(x, y, z)$ across a scale of 0.6 mm, 2.4 mm and 4.8 mm.

- $Gz_{LS}(x, y, z)$, the maximal value of the first order Gaussian derivative in the depth direction on the rib shadow likelihood image $LS(\mathbf{x})$ across a scale of 0.6 mm, 2.4 mm and 4.8 mm.

- $\frac{\sum_{k<z} LS(x,y,k)}{z}$, the average feature value (LS) of voxels above the voxel
- $\frac{\sum_{k>z} LS(x,y,k)}{N_z-z-1}$, the average feature value (LS) of voxels below the voxel
- $\frac{\sum_{k<z} LS(x,y,k)}{z} - \frac{\sum_{k>z} LS(x,y,k)}{N_z-z-1}$

We used a classifier with the same settings as the classifier used in the previous step except that the classifier in this step has a input layer of 9 nodes based on the features computed above. For the class of rib-surface points we use the manually annotated points. For the non-rib voxels we we used a set of regularly spaced samples that are at least 6 mm away from the chest wall (as defined by a fitted cylinder on the manually annotated rib-surface points).

This classification yields a likelihood for each voxel $LI(x, y, z)$, representing the probability that a voxel belongs to the rib surface (see Fig. 4.10).

The rib-surface points are determined by thresholding image $LI(x, y, z)$ with threshold defined as the 98th percentile of all likelihood values within the breast mask in each image.

Removal of False Positive due to the Nipple

Due to strong shadowing of the nipple and shadowing of ducts connected to the nipple, there were still a number of false positives were left after thresholding. Therefore among detected rib-surface points, points whose distance to the nipple in 2D coronal projection are within 9 mm are removed.

Cylinder Fitting to Detected Rib-Surface Points

In this step, we fit our cylinder model to the detected rib-surface points in a similar manner as described in Sec. 4.3.2. However, this time we weight the contribution of each rib-surface point by its likelihood LI:

$$\frac{1}{N}\sum_{n=1}^{n=N} w_n(\sqrt{(x_n - x_a)^2 + (y_n - y_a)^2 + (z_n - z_a)^2} - R)^2 \tag{4.13}$$

where w_n is defined as

$$w_n = LI(x_n, y_n, z_n)^{p1} * (\frac{d_{npl}(x_n, y_n)}{d_{npl_{max}}})^{p2} \tag{4.14}$$

$p1$ is chosen to enhance the weight difference of rib-surface points with different likelihoods and $p2$ is is chosen to enhance the weight difference of rib-surface points with different distance to the nipple and $d_{npl_{max}}$ is the maximal distance of the rib-surface points to the nipple ($p1$ and $p2$ were both experimentally set to 3).

Classifier Training

Given a dataset, we split it into two folds (fold 1 and fold 2) to perform two-fold cross-validation. To avoid bias, the two classifiers (rib-shadow classifier and rib-surface classifier) are both trained on data from fold 1 and tested on data from fold 2. To obtain the result for fold 1, we repeat the above procedure by swapping the two folds.

Implementation and Computation time

Our segmentation method was implemented in C++. Using a computer with a 2660 MHz CPU and 8 GB RAM, the total computation time for the chest wall segmentation in an image is about 6 minutes and 30 seconds in which the cylinder fitting procedure takes 1 minute and 30 seconds. We used a workstation developed by our group for visualization.

4.4 Results

We compute an error measure for each chest wall segmentation. This error measure is defined as the average distance of all annotated rib surface points to the automatically segmented chest wall. We applied our method separately on dataset A and dataset B and we also trained and tested our method on a combined dataset in which images of dataset A and dataset B were mixed (Table 4.1).

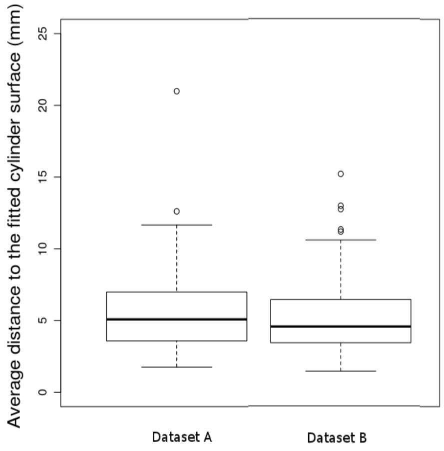


Figure 4.11: Box plot of the average distance of the annotated points (reference standard) in each image to the fitted cylinder surface using detected rib-surface points in dataset A and dataset B.

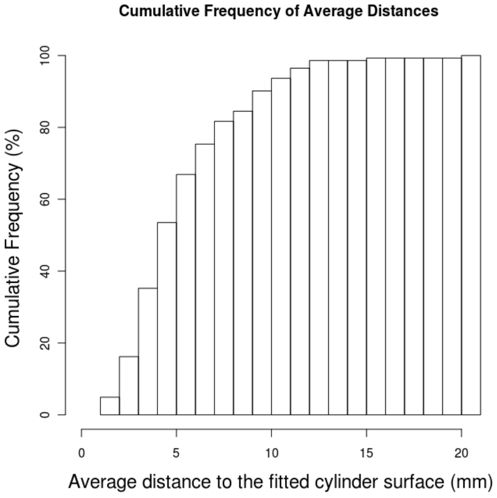


Figure 4.12: Cumulative histogram for the averaged distance of annotated rib-surface points to the segmented surface for all images in dataset A and dataset B with separated training.

(a) (b) (c)

Figure 4.13: Segmentation examples with chest wall segmentation as overlay. Three segmentation examples are shown in coronal and sagittal planes with an average distance between the automatically detected chest wall and the manually annotated rib surface points of 4.64 mm (a), 9.03 mm (b) and 12.61 mm (c) in ABUS, respectively. As we can see for case (c), our method over segmented the chest wall with a large error.

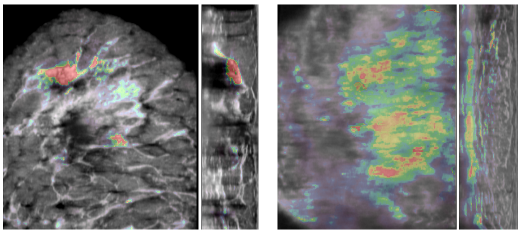
Table 4.1: Segmentation Results.

Testing set	Training set	Mean error
A	A	5.73 mm±3.11 mm
B	B	5.43 mm±3.07 mm
A + B	A and B separately	5.59 mm±3.08 mm
A + B	A + B pooled	5.95 mm±3.54 mm

For the 75 scans of dataset A, the mean error was 5.73 mm with standard deviation 3.11 mm. For the 63 scans of dataset B, the mean error was 5.43 mm with standard deviation 3.07 mm. For all 142 images, the mean error was 5.59 mm with standard deviation 3.08 mm. A boxplot and a cumulative histogram plot summarize the result (Fig. 4.11 and Fig. 4.12).

Three segmentation examples are shown in Fig. 4.13 with an average distance of 4.64 mm, 9.03 mm, and 12.61 mm, respectively. By visually inspecting our segmentation results we determined that an average distance of 8 mm or less generally results in an acceptable segmentation for the development of CADe systems and advanced visualization tools for ABUS reading. As seen from Fig.4.12, for all 142 images, about 82 % of scans had an average distance less or equal to 8 mm.

By pooling all 142 images together for training and testing, the mean error was 5.95 mm with standard deviation 3.54 mm which is significantly different to the resulting with separated training (paired t-test, p = 0.008).



(a) (b)

Figure 4.14: Two examples of cases with poor result in coronal and sagittal views with the likelihood map for rib surface as overlay. False positives are caused by dense tissue in case (a) and rib surface in case (b) are assigned low likelihood due to insufficient scanning depth.

4.5 Discussion and Conclusion

Our result in Section 4.3.2 demonstrate that the surface of the chest wall in an ABUS scan can be well approximated by the surface of a cylinder. Therefore we investigated an automated method to locate chest wall surface using this model. Using this the average of the mean distance of the annotated points to the segmented chest wall surface was 5.73 mm (dataset A) and 5.43 mm (dataset B), which is promising for follow-up processing, such as computer-aided detection and inter- and intra-modal image registration.

From the fitting error using manual annotations (2.19 mm for dataset A and 2.43 mm for dataset B, we can see there is a limitation of using the cylinder model. The limitation may be due to our cylinder model. To improve the method, in the future we can include additional parameters that allow bending of the cylinder as well as as an elliptical cross-section. We can also use our result as a starting point for further refinement, for instance, using graph search methods to optimize a defined layer or surface[34].

In our work, training and testing is involved for the evaluation. We investigated if it was necessary to use the images from the same manufactures for training and testing as opposed to mixing the images from two different manufactures. We found that by using images from the same manufactures for training and testing, significantly better results were obtained than by pooling images of different manufactures.

The segmentation method had poor results for a few cases. One reason for poor performance is that for extremely dense breasts, dense tissue also causes shadows behind it. Our rib-surface detection algorithm in those cases assigns high probabilities to the dense tissue, resulting in a poor segmentation of the chest wall. An example is shown in Fig. 4.14(a). Another reason is that for some images, rib shadows are not obvious enough due to insufficient scanning depth. For example in Fig. 4.14(b), the scanning depth is not sufficient to include rib shadows (the rib surface is too close to the bottom of image). For further improvement, false positives of the rib-surface detector must be reduced, for instance by incorporating more features characterizing dense tissue and ducts.

Acknowledgment

This work was presented at the MICCAI (the International Conference on Medical Image Computing and Computer Assisted Intervention) 2011 Workshop on Breast Image Analysis. This work has been funded by the HAMAM project(IST-2007-224538) within the Seventh Framework Programme (FP7) of the EU. The authors would like to thank André Grivegnée from Cancer Prevention and Screening Clinic, Jules Bordet Institute, Brussels, Belgium, László Tabár from Department of Mammography, Falun Central Hospital, Sweden and Matthieu Rutten from Department of Radiology, Jeroen Bosch Ziekenhuis, Den Bosch, the Netherlands for providing the data.

Computer-aided Detection in ABUS

5

Tao Tan, Bram Platel, Roel Mus, László Tabár, Ritse M. Mann, and Nico Karssemeijer

Original title: Computer-aided Detection of Cancer in Automated 3D Breast Ultrasound

Published in: IEEE Transactions on Medical Imaging 2013;32:1698-1706

Abstract

Automated 3D breast ultrasound (ABUS) has gained a lot of interest and may become widely used in screening of dense breasts, where sensitivity of mammography is poor. However, reading ABUS images is time consuming, and subtle abnormalities may be missed. Therefore, we are developing a computer aided detection (CAD) system to help reduce reading time and prevent errors.

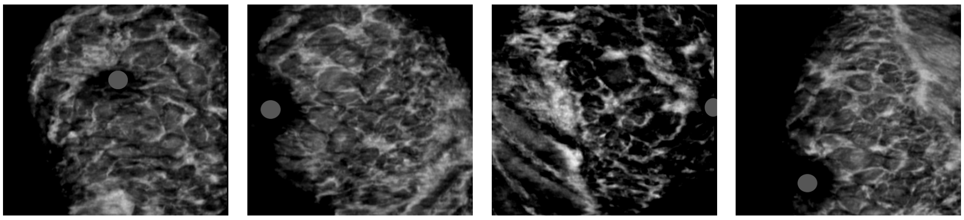
In the multi-stage system we propose, segmentations of the breast, the nipple and the chestwall are performed, providing landmarks for the detection algorithm. Subsequently, voxel features characterizing coronal spiculation patterns, blobness, contrast, and depth are extracted. Using an ensemble of neural-network classifiers, a likelihood map indicating potential abnormality is computed. Local maxima in the likelihood map are determined and form a set of candidates in each image. These candidates are further processed in a second detection stage, which includes region segmentation, feature extraction and a final classification. On region level, classification experiments were performed using different classifiers including an ensemble of neural networks, a support vector machine, a k-nearest neighbors, a linear discriminant, and a gentle boost classifier.

Performance was determined using a dataset of 238 patients with 348 images (views), including 169 malignant and 154 benign lesions. Using free response receiver operating characteristic (FROC) analysis, the system obtains a view-based sensitivity of 64% at 1 false positives per image using an ensemble of neural-network classifiers.

5.1 Introduction

Breast cancer is the leading cause of cancer death and the most frequently diagnosed cancer among women in both economically developed and developing countries[48]. One in every eight women will develop breast cancer in her life time. Studies show that screening reduces the breast cancer mortality rate[10,75]. Currently mammography is the primary modality for screening. However, the sensitivity of mammography is poor when breast density is high[7]. As an adjunct modality to mammography, whole breast evaluation with 2D hand-held ultrasound does increase screening sensitivity in dense breasts[6,53,55]. Use of targeted ultrasound in screening may reduce unnecessary recalls due to its ability to characterize lesions detected by mammography, in particular cysts[68]. However, 2D handheld ultrasound has operator-dependence and is time-consuming for whole breast evaluation. Moreover, it can not visualize breasts in 3D.

Automated 3D breast ultrasound (ABUS) was developed to alleviate the drawbacks of 2D ultrasound. The modality involves compression of the breast using a dedicated membrane and a wide transducer mounted in a scanning device. It provides 3D ultrasound images (views) of the breast from the skin line to the chest wall by an automated sweep over the breast. For each breast, depending on its size, up to as many as five views (anterior-posterior, medial, lateral, superior, and inferior views) are taken to cover the whole breast volume. Fig. 5.1 shows the most common four views for a selected breast. Different than with 2D ultrasound, with ABUS readers are able to inspect coronal slices of the breast. Spiculation or retraction patterns, often characteristic for malignancy are visible on these coronal slices, but are seldom seen on 2D ultrasound[83,106]. Wenkel et al.[107] show that using ABUS allows to achieve a high reliability of detection of solid and cystic lesions. Giuliano and Giuliano[37] showed that using ABUS in mammographically dense breasts could improve breast cancer detection in asymptomatic women. In a reader study[15], using ABUS led to a significantly higher sensitivity for malignant lesions than for benign lesions. Moon et al.[69] developed a computer-aided detection system for breast tumors in ABUS using multi-scale blob detection algorithm. In 2012, the U.S. food and drug administration (FDA) approved the use of an ABUS system developed by U-systems (Sunnyvale, CA, USA) for breast cancer screening.



(a) (b) (c) (d)

Figure 5.1: Coronal sections (views) of the right breast of (a) anterior-posterior, (b) medial, (c) lateral, (d) superior acquisitions. Each view covers a part of the breast and red circles indicate the nipple positions.

With ABUS systems the effort and costs of acquiring images can be reduced, since the imaging procedure can be performed by radiographers, as opposed to hand-held ultrasound which has to be performed by radiologists. Despite this, implementation of ABUS in screening will lead to a strong increase of the workload of radiologists, because reading ABUS screening exams is much more time-consuming than reading of mammograms. Apart from that, in the volumetric ABUS images lesions may be overlooked more easily than in mammograms. Therefore, there is a need to develop techniques to assist radiologists with the reading of ABUS images, to allow for more efficient reading while avoiding that lesions are overlooked. In screening mammography, computer-aided detection (CAD) systems have been considered as a tool to relieve the shortage of trained readers for double reading[4,76]. Moreover, different studies[9,36,72,86] have shown that use of CAD can improve the reader performance of breast cancer detection.

In this paper, we present a novel system for detecting breast cancer in ABUS, as an extension of previously reported work[96]. The method consists of an initial stage to locate lesion candidates by combining voxel features dedicated to detecting lesions such as blobness, spiculation and contrast and a second stage to eliminate false positive candidates by incorporating features characterizing the difference between cancers and non-lesion structures such as the shadows of nipple, ligaments and ducts. Evaluation was performed on a database of ABUS cases from 238 patients, acquired in four different centers with systems from two manufacturers. The evaluation of our CAD performance is based on free response receiver operating characteristic (FROC) analysis. It is noted that some of the features we use in this work are similar to those reported previously in a study on automated classification of benign and malignant lesions[94] as they have been proven to be very descriptive for cancers in ABUS images. The detection problem studied in this paper is quite different though from lesion classification, because here we mainly deal with classification between cancers and non-lesion false positives, which are different from benign lesions.

5.2 Materials and Methods

5.2.1 Dataset

The automated 3D breast ultrasound images used in this study are a representative sample of cases obtained in routine clinical care or screening from the Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands), the Jeroen Bosch Ziekenhuis (Den Bosch, The Netherlands), the Falun Central Hospital (Falun, Sweden), and the Jules Bordet Institute (Brussels, Belgium). The used images were generated by two types of ABUS systems: the SomoVu automated 3D breast ultrasound system developed by U-systems (Sunnyvale, CA, USA) and the ACUSON S2000 automated breast volume scanning system developed by Siemens (Erlangen, Germany).

In our dataset, images from the device by U-systems have a maximum size of 14.6 cm by 16.8 cm on the coronal plane and a maximum depth of 4.86 cm while images from the device by Siemens have a maximum size of 15.4 cm by 16.8 cm on the coronal plane and a maximum depth of 6 cm. The transducer of the U-systems device model used in this study had a fixed frequency of 8.0 MHz or 10.0 MHz while the frequency of the transducer by Siemens is variable between 5.0 and 14.0 MHz, and can be adjusted according to the breast size. Each 3D volumetric view by the device from U-systems

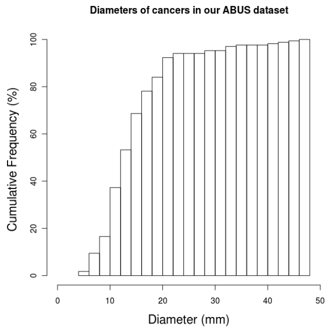


Figure 5.2: Cumulative histogram of lesion diameters of cancers in our dataset.

was generated with a minimal voxel size of 0.29 mm (along the transducer) by 0.13 mm (in depth direction) by 0.6 mm (along the sweeping direction) while images by the device from Siemens have a minimum voxel size of 0.21 mm by 0.07 mm by 0.52 mm. For processing, all images were resampled to obtain 0.6 mm cubic voxels.

The dataset used in this study consists of 348 images (views) from 238 patients (U-systems images: 51%, Siemens images: 49%). For patients with lesions, only those views in which the lesions were visible were included in this study. From these images, 169 malignant regions and 154 benign regions were annotated. From the 348 images, 40 images were completely normal. All malignant lesions were confirmed by biopsies. Benign lesions were either confirmed by biopsies or by information from images of different modalities. To develop and test our CAD system, we pooled all images and randomly split the dataset into two sets: Set A including 94 malignant and 69 benign lesions from 171 images and set B including 75 malignant and 85 benign lesions from 177 images. We took care that if there were multiple images of a patient these were not distributed over both sets to avoid bias. The diameter of annotated cancers in our dataset is 14.99 mm ± 6.99 mm. Fig. 5.2 shows the cumulative histogram for cancer diameters in our dataset. About 80 percent of our cancers have a diameter of less than 20 mm and over 50 percent cancers are with a diameter smaller than 15 mm.

5.2.2 Foreground Mask and Intensity Normalization

We first segment ABUS images to be able to restrict processing to the tissue volume. A foreground mask was generated by performing Otsu's thresholding[77] followed by a morphological closing operation to fill holes inside the mask[94]. It is noted that the foreground mask still includes tissue behind the chestwall and of the pectoral muscle. Ideally, this should be excluded too, but since no reliable segmentation method is available yet for ABUS images we apply the initial detection stage to the whole tissue

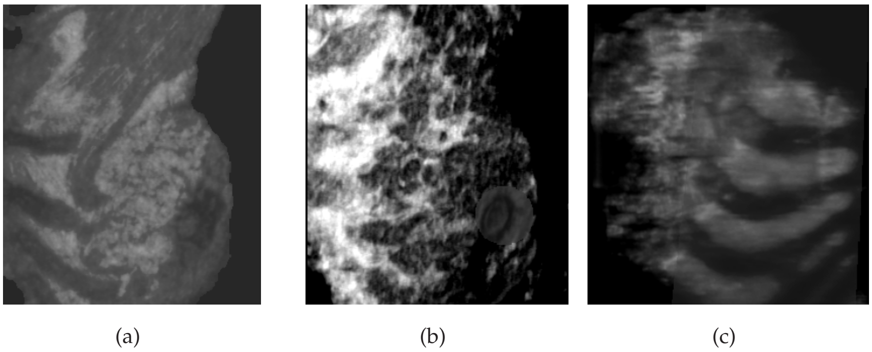

(a) (b) (c)

Figure 5.3: Coronal sections of (a) the foreground mask, (b) the nipple location, and (c) the chestwall segmentation.

mask.

Images in our dataset were acquired from two types of devices over a long period of time and in different institutes. Due to variations of scan parameter settings and use of different software versions the intensity levels of tissue types (fatty tissue, dense tissue, etc) varied from patient to patient. Therefore image intensities of each image were normalized. Since we only want to normalize the image based on the breast tissue, we exclude the top 6 mm slices to avoid responses of skin tissue and we exclude tissue with a depth more than 18 mm to minimize the presence of lung volume. Within the foreground mask, the $97\mathrm{th}$ percentile I'_{97} and the $5\mathrm{th}$ percentile I'_5 of intensity values of voxels with a depth between 6 mm and 18 mm were extracted. The normalized intensity I for each voxel is computed as:

$$I = \frac{(I_{97} - I_5)(I_o - I'_5)}{I'_{97} - I'_5} + I_5 \tag{5.1}$$

where I_o is the original voxel intensity and I_{97} and I_5 are constants representing reference values for the $97\mathrm{th}$ percentile and the $5\mathrm{th}$ percentile of intensity values.

The percentiles were chosen heuristically. We chose the $97\mathrm{th}$ percentile and the $5\mathrm{th}$ percentile of all intensity values as reference intensities for normalization instead of using the maximum and the minimum of all intensity values to avoid the influence of outliers on the normalization.

5.2.3 Chestwall Segmentation and Nipple Detection

After intensity normalization, a segmentation method is applied to find the chestwall[95]. Ideally, we want to exclude voxels behind the chestwall as CAD candidates. However, the chest segmentation is not accurate enough to retain all the lesions. Therefore we use the chestwall surface determined by the segmentation as a soft boundary. We do this by computing the signed distance to the surface, and by using this as a feature in the classification of lesion candidates. Pectoral muscle segmentation in ABUS has not been investigated in this study. Voxels in the pectoral muscles are included in the tissue evaluated by the system.

We observed that the shadows behind the nipple and the ducts connecting to the nipple are regions where potential false positives are caused. Therefore, we designed a method to detect the nipple and use it to compute a location feature. By including this in the lesion representation we can suppress false positives close or behind the nipple in the classification stage. The location of the nipple is automatically detected by using the Hough circle transform. This works well as the nipple in ABUS is visible as a dark circle in slices close to the transducer. Fig. 5.3 shows an example of foreground mask segmentation, nipple detection and chestwall segmentation.

5.2.4 Lesion Candidate Detection

In an initial lesion candidate detection stage, local features are calculated at each voxel across the image. Using these features, at each location an abnormality likelihood is computed using a classifier trained with labeled voxel data. Local maxima of the voxel likelihood map are then determined, and the highest maxima are taken to a next stage of the detection process.

We define a set of features to characterize lesion voxels in ABUS, representing blobness, coronal spiculation, contrast, and depth.

Blobness

As breast lesions appear as dark blob in ultrasound, we defined two features to characterize the blobness of the neighborhood surrounding a voxel. We applied an approach which was previously developed for the detection of masses in mammography[50,51]. For this work we extended the method to application in 3D.

In our method, in each voxel i, a measure of gradient convergence is computed which determines if there is a tendency of voxels within a certain distance of i to point to the central location i. To measure the convergence of neighboring gradients, we define a target sphere S_R with a radius R at voxel i and a spherical neighborhood as shown in Fig. 5.4. Voxels j with a distance r_{ij} ($r_{ij} \in [r_{min}, r_{max}]$) to i are selected when their gradient magnitude exceeds a small threshold. This selected set of voxels is denoted by N_i.

For any voxel j in N_i, a contribution x_j of voxel j to the convergence at site i is

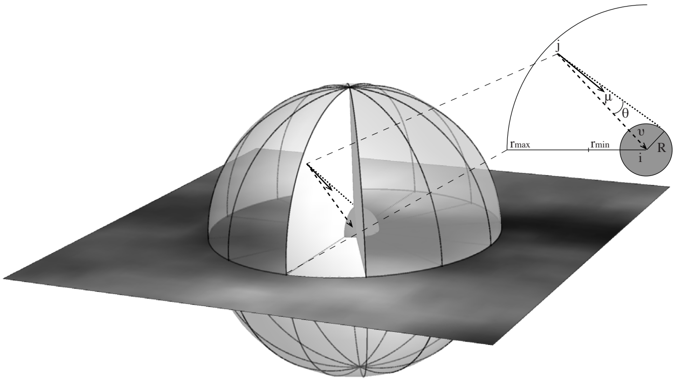


Figure 5.4: A schematic representation of the computation of our blobness feature. Voxels j that are located at a distance r_{ij} to the center voxel i between r_{min} and r_{max} and whose gradients are directed toward a sphere of radius R centered at the test site i make contributions to the blobness calculation.

defined as:

$$x_j = \begin{cases} 1 - p_j & \text{if } h(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\mu}}, \theta) = 1 \\ -p_j & \text{else} \end{cases} \tag{5.2}$$

$h(\nu, \mu, \theta)$ determines whether the gradient of voxel j is oriented towards the center i, which is defined as:

$$h(\nu, \mu, \theta) = \begin{cases} 1 & \text{if } \hat{\boldsymbol{\mu}} \cdot \hat{\boldsymbol{\nu}} > cos\theta \\ 0 & \text{else} \end{cases} \tag{5.3}$$

where $\hat{\boldsymbol{\mu}}$ is the unit vector of the gradient at voxel j, $\hat{\boldsymbol{\nu}}$ is the unit vector from j to i and θ is the angle between $\hat{\boldsymbol{\nu}}$ and a tangent line from j on S_R.

p_j is the probability that the gradient of voxel j would be pointing to the target sphere S_R at i in a random pattern with a uniform orientation distribution. To compute the probability we divide a partial area of a sphere, $S_{r_{ij}}$, with a radius r_{ij} centered at j by the whole area of the sphere $S_{r_{ij}}$. The partial area is the projection from the voxel j on the surface of $S_{r_{ij}}$ intersected by the target sphere S_R. Therefore, the probability p_j is

$$\frac{2\pi r_{ij}^2 (1 - cos(\theta))}{4\pi r_{ij}^2} = \frac{1}{2}(1 - cos\theta) \tag{5.4}$$

Using binomial statistics, the normalized blobness measure[50] at i is defined by

$$g1 = \frac{\sum_{j \in N_i} x_j}{\sqrt{\sum_{j \in N_i} p_j (1 - p_j)}} \tag{5.5}$$

To make this measure independent on the choice of the neighborhood size (r_{max}), it is computed as a function of r_{max} (4 mm $< r_{max} <$ 10 mm). The parameters r_{min} and R are both set to 2.4 mm. The maximum of $g1$ over the range of neighborhood size (r_{max}) was chosen as the blobness value of voxel i.

For the second blobness feature $g2$, when the maximum of $g1$ is chosen, we divide the spherical neighborhood into K_i sections as shown in Fig. 5.4. K_i was experimentally fixed to 40. This feature measures to what extent the gradients pointing to i are uniformly distributed among all sections. For a section k, the blobness contribution is

$$X_{i,k} = \sum_{j \in N_k} x_j \tag{5.6}$$

where N_k denotes voxels from N_i in section k.

The uniformity feature[50] is defined by

$$g2 = \frac{n_+ - K_i/2}{\sqrt{K_i/4}} \tag{5.7}$$

with n_+ the number of sections in which $X_{i,k}$ is positive.

Spiculation

Spiculation patterns are often seen in mammography when a cancer is present. Studies already showed that in the coronal plane of 3D ultrasound, spiculation patterns or retractions are often observed surrounding breast cancers[58,83,106]. Tan et al.[94] showed that coronal spiculation features contribute to automated discrimination of malignant from benign lesions. In this study, we used two spiculation features for voxel classification.

The method used to determine presence of spiculation was described in[50] and[94]. In this method, in each voxel i a measure of line convergence is computed which indicates if there is a pattern of white lines in the coronal plane pointing to i. We first define at location i on the coronal plane a circular neighborhood with radius r. Next, we analyze the voxel orientation pattern in this neighborhood, determined by matching at each site in the neighborhood the orientation of a directional 2nd-order Gaussian derivative with the local intensity pattern. When a spiculation pattern is present, voxels in this neighborhood are more likely to be oriented towards the central voxel i. We derive the following two features from the calculated voxel orientations. The first feature $sp1$ is a normalized measure of the fraction of voxels with orientation pointing towards the center. The second feature $sp2$ indicates whether voxels with orientations pointing to the center are uniformly distributed among all angular sections. Similar to the computation of the blobness features, the final output of $sp1$ is the maximum of the values that were compute at several scales (12 mm $< r <$ 30 mm) and $sp2$ is computed at the scale when the maximum of $sp1$ is found.

Contrast

To determine contrast of the region centered at a voxel i, we define a spherical neighborhood of i with a radius r. Two contrast features are computed as a function of r. The first is defined by

$$c1(r) = I_s(r) - I_s(r - d) \tag{5.8}$$

where $I_s(r)$ represents the average intensity measured in the spherical neighborhood of i. The parameter d was fixed to 3 mm empirically.

The second contrast feature was computed similarly but only in 2D coronal planes, using a circular neighborhood around i with a radius r. This feature was defined by

$$c2(r) = I_c(r) - I_c(r - d) \tag{5.9}$$

where $I_c(r)$ represents the average intensity measured in the circular neighborhood. The parameter d was fixed to 3 mm. The computation of the final features $c1$ and $c2$ is similar to that of the spiculation and blobness features. The final output of $c1$ and $c2$ is the maximum of the values computed in the range (4 mm $< r <$ 10 mm).

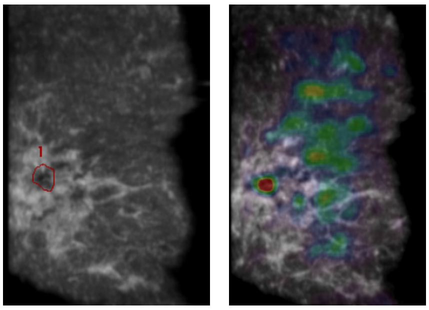
Depth

In ultrasound features change with increasing depth. In addition, the probability of occurrence of cancers also varies with depth. To take this into account, we used the depth of voxels dp as a feature. The idea is that this will help to reduce false positives (FP) which are very deep or very close to the skin.

5.2.5 Voxel classification and lesion candidates

To obtain lesion candidates, an initial voxel classification is performed using the seven features described above. Supervised classification was used with an ensemble of five neural networks. Every neural network consists of an input layer of 7 nodes representing the features, and a hidden layer of 5 hidden nodes and an output layer with one node. Each network was initiated with random weights to reduce the influence of the initial weights on the final output of the trained network. The final classifier output was computed by averaging the five network outputs, which stabilizes classification outputs and improves the classification performance compared to a single neural network classifier as suggested by studies[39 108].

Training data for the voxel classifier was selected using the ground truth annotations. For the initial detection stage, we trained our system to detect both benign and malignant lesions. Given that the average breast lesion radius is 9 mm[100] and the average cancer radius is 7.5 mm in our dataset, for the class of lesion voxels, we empirically



(a) (b)

Figure 5.5: Coronal sections of an anterior-posterior of the right breast acquisition (a) with an annotated malignant lesion, (b) overlayed with the abnormality likelihood map, where red represents areas of high likelihood.

used all voxels within 3.6 mm (6 voxels) to the center of annotations of benign and malignant lesions. For the class of normal voxels, we selected voxels within the breast mask that were located at least 18 mm away from the lesion center.

An example of a likelihood map obtained from the voxel classifier is shown in Fig. 5.5. Subsequently the likelihood map is smoothed using a Gaussian kernel, thresholded, and local maxima are computed. Maximum within 10 mm of each other are fused into one new maximum. At most 16 maxima per view were kept from the initial detection. The resulting maxima are then passed on to the second stage of the detection.

5.2.6 False positive reduction

The lesion candidate locations obtained from the initial stage of the detection are used as seed points for a 3D region segmentation. We use a spiral-scanning based dynamic programming technique, which was initially introduced by Wang et al.[104] for pulmonary nodule segmentation in CT. The method was adapted by Tan et al.[94] for lesion segmentation in ABUS. Subsequently, region features are extracted to train a classifier for false positive reduction. The features used are described below.

Features from initial stage

Given a segmentation of a region, the output of the voxel based classifier in the initial stage and the voxel features can be summarized in many ways to represent a region. Besides the voxel features and the voxel likelihood of the seed point, we use the mean and the maximum value of $g1$, $g2$, $sp1$, $sp2$, $c1$, and $c2$ inside the segmented region. In addition, the $75th$ percentile of $sp1$, $sp2$ and the voxel likelihood l in the region were also extracted. Furthermore, we use the difference between the voxel likelihood of the seed point and the median of the likelihoods of all seed points from the same image, to indicate the relative abnormality of this seed point. Similarly, the difference between the intensity of the seed point and the median of intensities of all seed points was also used to indicate the relative darkness of the candidate region. In total, this group comprised of 25 features.

Texture

Within a minimal cubic bounding box surrounding a segmented candidate region, the co-occurrence matrix with an offset of 3.6 mm in the depth direction is used to derive 6 texture features[40] including energy, entropy, inverse difference moment, inertia, cluster shade and cluster prominence.

Location

Shadows and structures caused by the nipple and the connecting ducts very close to the nipple are often assigned high likelihoods in the initial stage. Therefore, the relative location of a candidate region to the nipple is important. These effects are mostly limited to 6 cm wide area around the nipple in coronal plane. We include a feature describing relative location. For regions closer than 3 cm to the nipple in the coronal projection, the distance to the nipple was used as a feature. For regions further away, a feature value of 3 cm was used, thus avoiding that this influences discrimination of regions further away from the nipple. Furthermore, to reduce false positives beyond the chestwall, a signed distance to the chestwall was included in the feature set. The distance to the boundary of the foreground mask, seed point depth, and the depth of both the highest and lowest voxel of the region segmentation were also used. Thus, 6 features related to lesion location were used.

Dedicated Region Features

To represent other relevant properties of the segmented regions we compute volumetric height-to-width ratio, posterior acoustic behavior, compactness, average region intensity, region volume[96]. Furthermore we compute the difference between the average intensity of a region and its outer border, consisting of voxels outside the lesion within 1.2 mm of the boundary, and the average radial gradient magnitude of the region boundary. Margin contrast is computed in a central segment of a region.

True lesions in ABUS appear darker than fatty and dense tissue. Dark tissue also includes shadows and artefacts. To reduce false positives in fatty and dense tissue

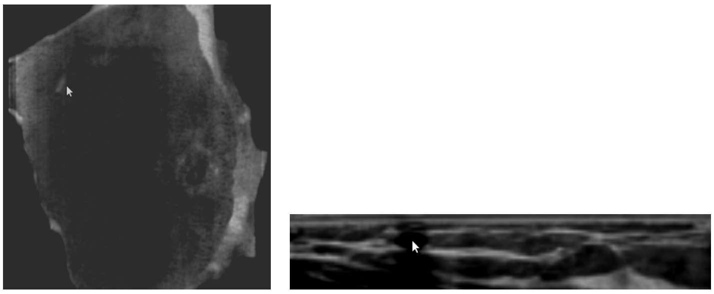


Figure 5.6: The topmost coronal section of an anterior-posterior acquisition of a breast showing (a) an area of insufficient contrast with the transducer, (b) shadows caused by the lack of contract.

regions, we compute the fraction of dense and fatty tissue in a segmented candidate region. For this purpose we segment each ABUS image in fatty, dense and dark tissue using Otsu's thresholding[77].

During ABUS imaging, bad skin contact of the transducer, often due to insufficient use of contact lotion, results in artefacts. These appear as bright blobs in the upper coronal slices with a strong shadowing effect in the lower slices (Fig. 5.6). To avoid false positive due to these artefacts, we project the region segmentation on the upper coronal slice. The average intensity of the projected segmentation on the upper coronal slice, I_{Top}, and the difference between the average intensity in projection and its outer border, consisting of voxels outside the lesion within 3 mm of the boundary, $I_{TopContrast}$ were computed to characterize these false positive regions.

The first stage classification also often assigns a high response to thin elongated shadow patterns caused by ducts or ligaments only visible in few transversal or sagittal slices. Since small compact tumors are often round we include an area ratio as a feature defined by

$$ar = \frac{|area_t - area_s|}{\max(area_t, area_s)} \tag{5.10}$$

where $area_t$ and $area_s$ are region size on transversal and sagittal views through the center, respectively.

5.2.7 Region Classification

In the second stage of the detection, a suspiciousness level for a region being a cancer is computed by the CAD system using the features described above and supervised classification. In total 50 features were included. We investigated the use of five different classification methods: A support vector machine (SVM) with a radial basis function (RBF), an ensemble of five neural networks (NN), linear discriminant analysis (LDA), gentleboost (GB)[31,102], and k-nearest neighbors (kNN). To train the region classifier, we took candidate regions whose centers were closer than 10 mm to the annotated malignant lesion center for the class of cancer regions. For the class of normal regions we took the candidate regions whose centers were more than 10 mm away from both malignant and benign lesion centers.

5.2.8 Experiments and Evaluation

To obtain unbiased detection results we used the following procedure. In the initial stage, we used Set A for training and Set B for testing. In the second stage, a 10-fold cross-validation procedure on patient level was performed on the testing dataset Set B. For each run of the cross-validation, the training folds from Set B and the whole Set A were combined into a training set. This training set was also used to perform selection of the parameters of the classifiers using a grid search process, such as the penalty factor and kernel parameter for SVM, the number of stumps for GB, the fraction of abnormal samples for NN, and the number of neighbors for kNN. The outer loop of the cross validation was used to determine system performance. To obtain an unbiased detection result for Set A, we repeated the same procedure by swapping the two subsets. The detection results of Set A and Set B were pooled to estimate the performance of the CAD system.

To investigate the benefits of using the spiculation features in the initial detection stage, we conducted experiments with and without using the spiculation features.

For both stages of the detection we use free-response receiver operating characteristic (FROC) analysis to determine the performance of the system based on the suspiciousness level obtained for each region or voxel candidate (seed point) and ground truth annotations. The annotations we had available were 2D contours on coronal slices, representing the sections where the lesions are best visible. A region or voxel candidate is considered a hit when the center of the region or the seed point is within 10 mm to the annotated malignant lesion center. Each annotation can only obtain one hit. When multiple regions or seed points hit the annotation of a malignant lesion, the one with the highest suspiciousness is chosen and the rest is discarded. The hits on annotated benign lesions are ignored and thus not considered as false positives.

We used view-based and patient-based analysis. This leads to different results when lesions are visible in multiple views, or if more than one region associated with cancer is present in a view. Using view-based analysis all regions in different views are treated independently. The positives are annotated cancer regions and true positives are correctly detected cancer regions. With patient-based analysis the true positive fraction is computed as the fraction of cancerous patients in which at least one cancerous region is detected out of all cancerous patients. When cancers are detected in different views of the same patient, the CAD region with the highest likelihood is assigned to the patient. The false positives in different views are always counted separately.

For statistical analysis of the results of our experiments we used bootstrapping. For each experiment, the test dataset was sampled with replacement 1000 times. As a measure of performance we used the difference of the mean sensitivity of the system in the range between 0.1 FP/image and 10 FPs/image, computed on a logarithmic scale. To determine significance of experimental outcomes p values were obtained by taking the fraction of positive or negative differences in the distribution. A difference in performance of two classifiers is considered to be significant if $p < 0.025/N$ (two-sided), with N the number of comparisons performed for Bonferroni correction.

5.3 Results

Fig. 5.7 shows the performance of the initial detection and second detection stage of the system, using the neural-network classifiers in both stages. In the initial stage, the CAD system achieved a view-based lesion sensitivity of 48% and 56% at one and two false positives per image, respectively. About 12% of the malignant lesions is missed by the initial stage. With the second stage, view-based lesion sensitivity increased to 64% and 70% at one and two false positives per image, respectively. Fig. 5.8 shows an example of a cancer that is correcly detected by the system. Table 5.1 and Fig. 5.9 summarize the performances of the second stage using different classifiers. At one false positive per image the NN classifier had the highest view-based lesion sensitivity (64%). Fig. 5.10 shows the performance of the NN based system using view-based and patient-based lesion sensitivity. At one false positive per image patient-based lesion sensitivity was 73%. Fig. 5.11 shows that by using spiculation in the initial detection stage, the performance is significantly improved ($p < 0.001$).

Table 5.1: The sensitivities of the experiments using different classifiers at 1 FP/image and 2 FPs/image.

	NN	SVM	kNN	LDA	GB
1 FP/image	64%	62%	61%	62%	57%
2 FP/image	70%	68%	68%	68%	60%

Table 5.2 shows p values for the comparisons between second stage classifiers using view-based analysis. Using the neural-network classifier, the performance is signifi-

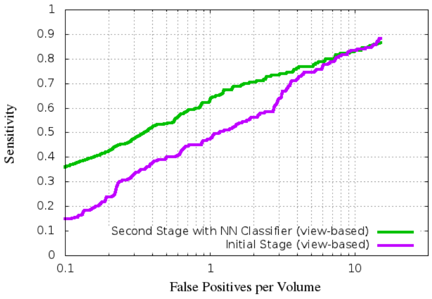


Figure 5.7: FROC curves of the initial stage and the second stage of the CAD system using the neural-network classifier.

cantly better compared to that using GB ($p < 0.025/10$, considering that 10 comparisons between different classifiers are performed).

Table 5.2: The p values between the performances by using different classifiers in the second stage. p values are bolded when the mean sensitivity using classifier in the left entry is higher than that in the upper entry.

Classifiers	NN	SVM	kNN	LDA	GB
NN	-	**0.330**	**0.031**	**0.009**	<**0.001**
SVM	-	-	**0.072**	**0.019**	<**0.001**
kNN	-	-	-	**0.313**	**0.003**
LDA	-	-	-	-	**0.054**
GB	-	-	-	-	-

5.4 Conclusion and Discussion

We have developed a multi-stage computer-aided detection (CAD) system for breast cancers in automated 3D breast ultrasound (ABUS). It was found that by using a second detection stage in which a region based classification is performed based on dedicated features we developed to reduce false positives the performance of the initial detection is improved. Using a large databse for evaluation, we found that at one false positive per image the view-based lesion sensitivity and the patient-based lesion sensitivity were 64% and 73%, respectively. Whether this is sufficient for clinical application is not yet known, since no systems for detection of lesions in ABUS have been clinically evaluated to date.

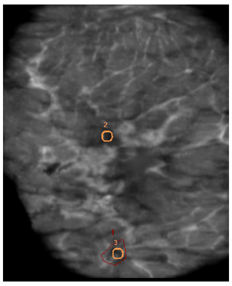


Figure 5.8: Coronal section of an anterior-posterior acquisition of a breast showing the CAD regions (orange circles) from the second stage detection and the corresponding cancer annotation (red contour).

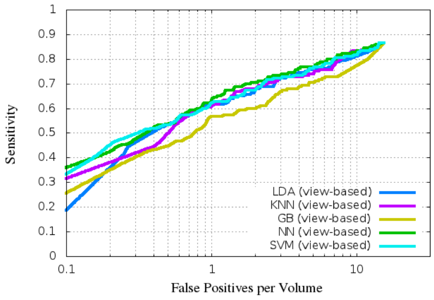


Figure 5.9: FROC curves of the second stage of the CAD system using different classifiers.

In the initial detection stage of our method, seven features were used for voxel classification. We found that spiculation features were particularly effective. The performance of a system using coronal spiculation features was significantly better than that of a system without these features.

The aim of the initial detection is to detect as many true lesions as possible at a false positive rate that is still manageable in second stage. In the design of our system, cancers missed by the first stage cannot be detected in the second stage, unless the region segmentation from other lesion candidates accidentally extends to the undetected cancers. We found that at the cutoff point we used twelve percent of malignant lesions were missed by the initial detection. Fig. 5.12 shows an example of such a missed lesion. In this case the lesion was missed because there is no spiculation or retraction pattern present and because the lesion shape is not typical for cancer. Without a spiculation or blobness signal these lesions are missed by the current system. In the future, voxel features characterizing lesions with a less blob-like shapes need to incorporated. Some other cancers missed by our system appeared to be rather large, with a radius

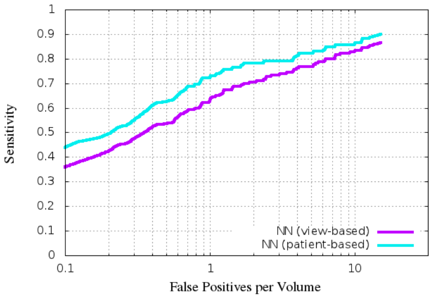


Figure 5.10: View-based and patient-based FROC curves of the second stage of the CAD system using an ensemble of five neural-network classifiers.

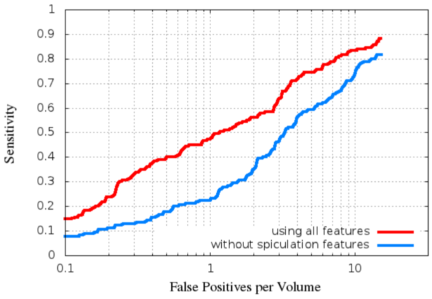
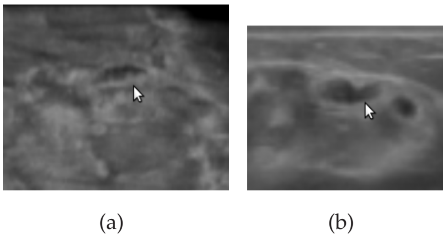


Figure 5.11: FROC curves of the initial stage with and without incorporating spiculation features using an ensemble of five neural-network classifiers.

exceeding 20 mm. In fact, it turned out that some were detected with a mark in the periphery of the lesion, but not counted as true positive due to our detection criterium using a distance measure. Since such large lesions are not well represented in the training data they are easily missed, but one might argue that missing these lesions would not have a large implication in practice because radiologists will not overlook them.

For the region classification, a number of classifiers were used and the resulting performances were compared. Using an ensemble of neural-network classifiers, the performance was significantly better than when using the gentleboost classifier. This was surprising since we expected better performance with the boosting techniques, which are often found to have superior performance in the literature.

The performance of our CAD system is evaluated by FROC analysis. If the system is used clinically to mark suspicious regions, an operating point has to be chosen at which the number of false positives is acceptable for radiologists. In mammography commercial CAD systems currently have about 1 to 3 marks per case[60,62]. Because



(a) (b)

Figure 5.12: A malignant lesion missed by the initial detection is shown in (a) coronal and (b) transversal plane.

we are dealing with volumetric data, more markers might be acceptable. While an appropriate setting has to be determined experimentally, we believe that more than two marks per volume will not be tolerated in practice. At that point, our system has a patient-based sensitivity of 78%. Future research should improve performance and both reduce false positives and increase sensitivity. Furthermore the detection system can be used in combination with a dedicated computer-aided classification system to distinguish malignant lesions from benign lesions.

Acknowledgment

The authors would like to thank André Grivegnée from Cancer Prevention and Screening Clinic, Jules Bordet Institute, Brussels, Belgium, and Matthieu Rutten from Department of Radiology, Jeroen Bosch Ziekenhuis, Den Bosch, the Netherlands for providing data for this study.

Finding Lesion Correspondences in Different Views of Automated 3D Breast Ultrasound

6

Tao Tan, Bram Platel, Michael Hicks, Ritse M. Mann, and Nico Karssemeijer

Original title: Finding Lesion Correspondences in Different Views of Automated 3D Breast Ultrasound

Published in: Proc. SPIE 8670, Medical Imaging 2013

ABSTRACT

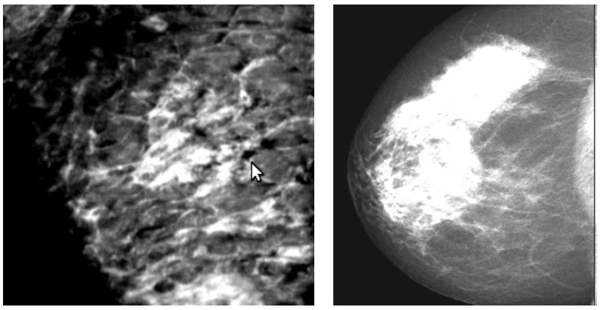
Screening with automated 3D breast ultrasound (ABUS) is gaining popularity. However, the acquisition of multiple views required to cover an entire breast makes radiologic reading time-consuming. Linking lesions across views can facilitate the reading process. In this paper, we propose a method to automatically predict the position of a lesion in the target ABUS views, given the location of the lesion in a source ABUS view. We combine features describing the lesion location with respect to the nipple, the transducer and the chestwall, with features describing lesion properties such as intensity, spiculation, blobness, contrast and lesion likelihood. By using a grid search strategy, the location of the lesion was predicted in the target view. Our method achieved an error of 15.64 mm±16.13 mm. The error is small enough to help locate the lesion with minor additional interaction.

Keywords: automated 3D breast ultrasound, breast cancer, CAD, automated linkage

6.1 Introduction

It is well known that the sensitivity of mammography is low in dense breasts. In mammograms of these women, small masses can be obscured by the dense fibroglandular tissue (Fig. 6.1)[59]. Schaefer et al.[89] showed that by using handheld ultrasound together with mammography, 15.9 % extra cancers were found for the dense breasts and 12.8 % extra cancers were found for all breasts. Berg et al.[6] showed that using ultrasound and mammography could reveal additional 1.1-7.2 cancers per 1000 high-risk women , with an increase of false positives. Automated 3D breast ultrasound (ABUS) is rapidly gaining popularity as a low cost adjunct screening modality for women who have dense breast tissue. Wenkel et al.[107] show that using ABUS allows to achieve a high reliability of detection of solid and cystic lesions. Giuliano and Giuliano[37] found that using automated 3D breast ultrasound (ABUS) together with mammography improved breast cancer detection rate, resulting 12.3 per 1000, compared to 4.6 per 1000 by mammography alone. In a reader study[15] by Chang et al., using ABUS led to a significantly higher sensitivity for malignant lesions than for benign lesions. In 2012, the U.S. food and drug administration (FDA) approved the use of an ABUS system developed by U-systems (Sunnyvale, CA, USA) for breast cancer screening.

Depending on the breast size, usually up to as many as five 3D ABUS acquisitions (called views, Fig. 6.2) are taken per-breast, in order to image the entire volume. Radiological reading of such a large amount of image data is time consuming. Existing computer-aided detection (CAD) methods for ABUS[96] can detect potential cancer



(a) (b)

Figure 6.1: (a) ABUS image containing shows a small cancer with a diameter < 6 mm indicated by the cursor (coronal). (b) The corresponding mammogram in CC view in which the lesion is obscured by dense tissue.

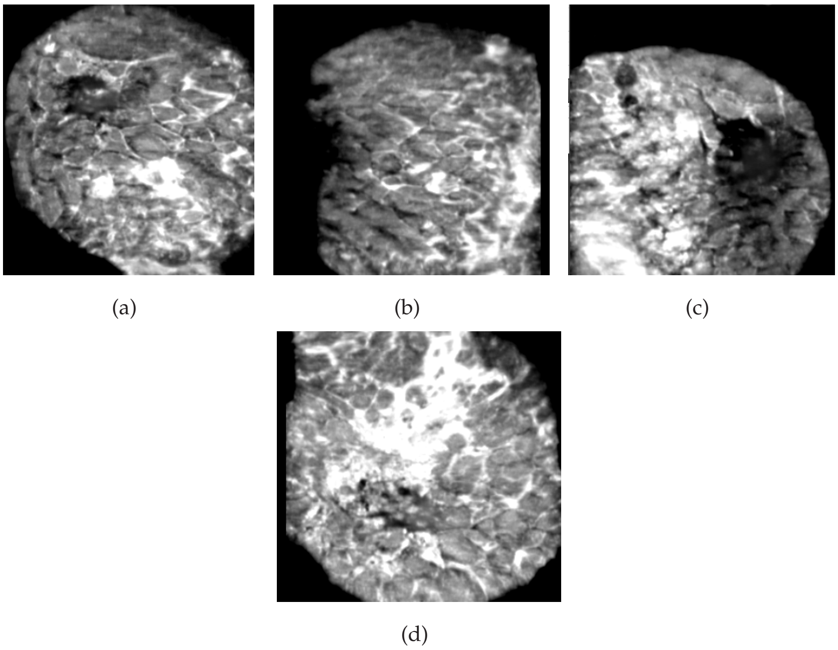
regions and their suspiciousness levels to facilitate the reading process. These algorithms, however, do not link the detected regions across the different views of the same breast. To improve localization and detection performance of readers, and to facilitate the further use of the CAD, we propose a method that allows radiologists to locate potential abnormalities in other views (target images) when a lesion has already been pointed out by reader or CAD in the current view (source image). This method can also be used to provide the linkage of CAD regions in different views to improve overall CAD performance.

6.2 Materials and Methods

6.2.1 Dataset

The breast ultrasound images used in this study were obtained from the Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands) and the Jules Bordet Institute (Brussels, Belgium). Both centers used the Siemens ACUSON S2000 ABVS (Erlangen, Germany). Imaging involved a frontal compression by a dedicated membrane and a transducer (5.0 to 14.0 MHz) with a width of 15.4 cm. Each 3D volumetric view was generated with a minimal voxel size of 0.21 mm (along the transducer) by 0.07 mm (in depth direction) by 0.52 mm (along the sweeping direction). For processing, images were resampled to obtain 0.6 mm cubic voxels.

For this study, we used a dataset containing ABUS images of 44 patients in which a lesion is visible in at least two views of the same breast. Corresponding lesions were annotated and manually linked as reference standard. Given the annotation of a lesion in one view, the task is to predict the position of the same lesion in the other view(s). We



(a) (b) (c)

(d)

Figure 6.2: Coronal slices of the four most common views in ABUS: (a) anterior-posterior view, (b) medial view, (c) lateral view, and (d) superior view.

selected only the views in which lesions are visible. In total, this yielded 120 directional linkages between the lesions.

6.2.2 Methods

A general approach to obtain correspondences between two images of the same object is by image registration. However, for this particular task, direct image registration is challenging. First, different ABUS views of the same breasts cover different parts of the breast and the small overlap and different coverages of different views would make registration less reliable. Secondly, due to large compressions in different directions during imaging, there is a large deformation between different views of the breast.

In this paper, we do not intend to find all voxel correspondences between different views. We are only interested in finding lesion correspondences. Therefore, we do not use an image-registration approach. Instead, we try to find the point in the target image which is closest to the source point in a defined feature space.

In order to restrict the predicted location to the breast tissue, segmentation of the breasts is necessary. A 3D breast mask was generated by performing Otsu's thresholding followed by a morphological closing operation to fill holes inside the mask[94].

When radiologists look for corresponding lesions in different views of the ABUS or between images from different modalities, they relate the position of the lesion to visible anatomical structures such as the skin of the breast, the nipple and the chestwall. To extract this information, the location of the nipple is automatically detected by using the Hough Circle Transform. This works well as the nipple in ABUS is visible as a dark circle in slices close to the transducer. The nipple location is used to compute contextual features for our algorithm. Furthermore, the chestwall is segmented using a cylinder model[95]. Fig. 6.3 shows an example of the breast mask segmentation, nipple detection and chestwall segmentation.

For each voxel within the breast mask, we compute position features using these landmarks and a set of local features to characterize lesions to find the corresponding lesions in two views.

Position Features

Given the location of the nipple in the 2D coronal plane and the chestwall segmentation, for each voxel we can compute its coordinate x, y relative to the nipple position (x is in the direction parallel to the transducer and y is in the direction of the transducer motion), the distance to the nipple dn, and signed distance to the chestwall dc. We also incorporated depth (distance from the membrane, dp) as a position feature.

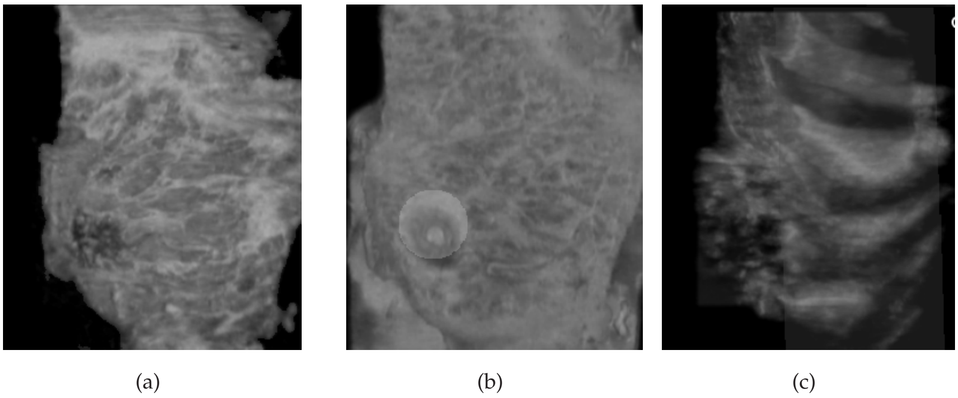

(a) (b) (c)

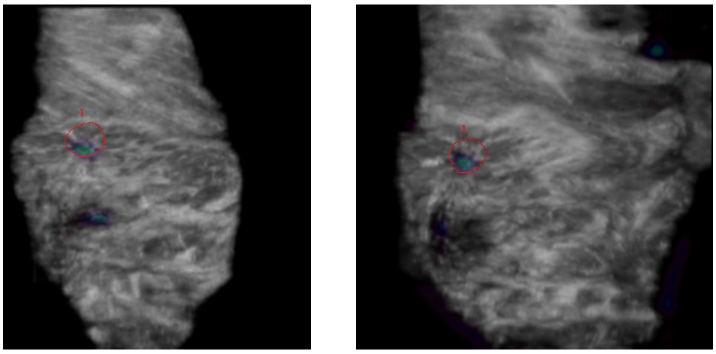
Figure 6.3: (a) a breast mask overlay in the coronal view, (b) nipple detection in the coronal view, (c) chestwall segmentation as overlay.

Local Features

Each view is obtained by compression of the breast towards the chest under different orientations. Therefore, there is a large deformation between different views. Due to this deformation, the predictions would be inaccurate if linkage would be based solely on the nipple and chestwall locations. This is why besides using the location information obtained from the segmentation step, for each voxel we extracted intensity I, two spiculation features ($sp1$, $sp2$), two blobness features ($b1$, $b2$), two contrast features ($c1$,$c2$), and the likelihood of a voxel being part of a lesion (l) (Fig. 6.4) resulting from a CAD system[96].

Location Prediction

The prediction of the lesion location is based on similarities of computed features between voxels of the lesion in the source image and voxels in the target image. All features are normalized such that for the view at hand the distribution of each feature has a zero mean and unit standard deviation using all voxels in the source view within the breast mask. We used the center of the annotation and its 6 neighbors from the source image as source points. In the target image, we use a grid search that compares the Euclidean distance of each voxel to the source point in the feature space. The matching voxel in the target image is the voxel with the smallest Euclidean distance to the source point. This results in 7 matching locations in the target image. The final location of the corresponding lesion is determined by the median location of these 7 points.



(a) (b)

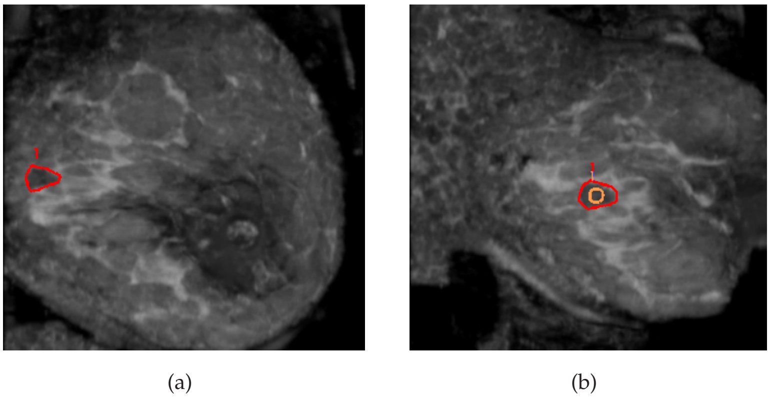
Figure 6.4: Coronal sections of a right breast with a lesion (red contour) and lesion likelihood overlay in (a) anterior-posterior view and (b) medial view, where green represents areas of high likelihood.

6.3 Results

We used the spatial distance in mm's of the predicted location to the center of the annotated location of the target lesion as our performance measure. We analyzed this measure for all linked lesions in our dataset bidirectionally. Fig. 6.5 shows an example of such a prediction. We perform our method with and without local features. Fig. 6.6 shows the box plots of errors obtained by the two experiments. With only position features, our method obtained a mean error of 20.28 mm and a standard deviation of 13.42 mm. With all features included, our method obtained a mean error of 15.64 mm and a standard deviation of 16.13 mm. Fig. 6.7 also shows the cumulative histogram of the errors obtained by the two experiments. Using all features, it can be seen that about 73% of the predicted locations are less than 20 mm away from the actual lesion center. Adding local features to the position features significantly improved the localization (paired t-test, p = 0.02).

6.4 Conclusion and Discussion

In this paper, we have presented a method to predict the position of a breast lesion in target ABUS views given the location of the lesion in a source ABUS view. We combined features describing the lesion location with respect to landmarks such as the nipple, the transducer and the chestwall, with features describing lesion properties such as intensity, spiculation, blobness, contrast, and likelihood of being a lesion. By using a grid search strategy, the location of the lesion was predicted in the target view.

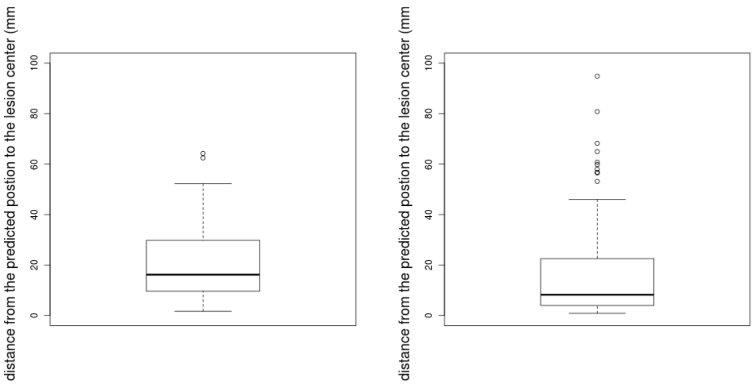


(a) (b)

Figure 6.5: (a) a lesion (red contour) in a source image. (b) the predicted position (orange circle) in a target image of the same breast.

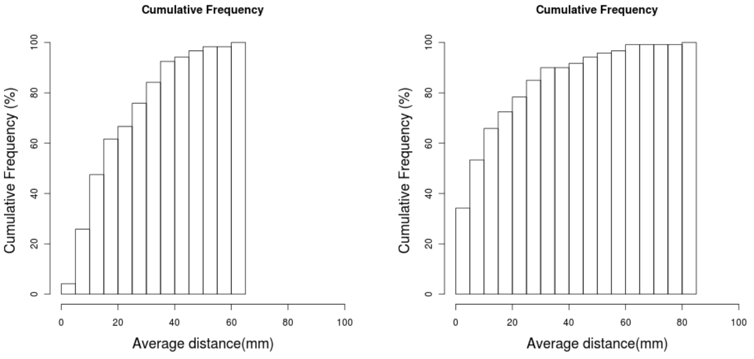
Our method achieved an error of 15.64 mm±16.13 mm. Considering that the average breast lesion radius is 9 mm[100], the error is likely small enough to help locate the lesion with minor additional interaction.

The preliminary results are promising. The accuracy of the proposed method mainly depends on the accuracy of the nipple detection. Using other features characterizing lesions significantly improved the accuracy. In the future, we will improve the method by incorporating features extracted at the region level, i.e. the size of the region, acoustic posterior behavior, etc. With those region features, a further improvement is expected, and by using more advanced prediction methods using supervised learning we could weight the influence of the different features on the result. Moreover, using a biomechanical model[65] of a breast to simulate the compressions of different views might also be helpful for the task.



(a) (b)

Figure 6.6: (a) boxplot of errors using only position feature vector. (b) boxplot of errors using all features.



(a) (b)

Figure 6.7: (a) cumulative histogram of errors using only postion feature vector. (b) cumulative histogram of errors using all features.

Summary and Discussion

Summary and Discussion

About 40 percent of women have dense breast tissue[8]. Women with extreme dense breasts have a risk of developing breast cancer that is 4 to 6 times higher than that of women of similar age with no or very low breast density[7]. For these dense breasts, the sensitivity of using mammography for cancer detection is as poor as 30-48%[63]. For female BRCA1/2 gene mutation carriers, exposure to diagnostic radiation before age 30 was associated with an increased risk of breast cancer[78]. For these women with different breast density and gene mutation, choosing a personalized screening modality is rather important.

Digital breast tomosynthesis mammography (DBT) as a promising new technology is developed for imaging dense breast. A 3D volume of a breast can be reconstructed using multiple 2D projection images. Studies has shown that using DBT together with mammography, the recall rate can be reduced[38] and the sensitivity can also be improved[3]. However, this technology is rather new, a fair estimate of the benefits applying DBM on personalized screening should be investigated. Moreover, the dose use of DBT is even higher than that of mammography. Therefore it is not suitable to apply DBT on BRCA1/2 gene mutation carriers. In the other hand, screening studies already showed that combining 2D ultrasound[6,12] or 3D[37,53] ultrasound together with mammography, the detection rate of breast cancers was significantly improved. Automated 3D breast ultrasound has a potential to be used as an alternative screening modality for these women with dense breasts and also for BRCA1/2 gene mutation carriers. The standardized imaging protocol make it more suitable as a screening device compared to 2D ultrasound.

To be able to image the entire breast, three to five ABUS views targeting different areas are taken to cover the whole volume of the breast while for mammography usually only 2 (MLO and CC) mammographic views are taken. Each ABUS views is an volumetric image consisting of around 280 transversal slices while a mammographic view is a single 2D slice. Therefore ABUS screening requires more man power for radiological reading compared to mammography screening. To speed up the workflow and make the ABUS screening more effective, computer systems are expected to play a role in the future. A computer aided detection (CADe) system can be used to facilitate the localization of suspicious regions and prevent oversight errors by radiologists. A computer-aided diagnosis (CADx) system can be used to distinguish malignant lesions from benign lesions, thus reducing the number of benign lesion recalls.

Chapter 2 reports on the development of a computer-aided diagnosis (CADx) system for characterization of breast lesions in ABUS. In our work, we developed a system that extract features describing spiculation from reconstructed coronal planes. In this plane spiculation is better visible than in other planes and we know this is highly correlated with malignancy. The spiculation feature is absent in the most CADx systems developed for 2D ultrasound and 3D ultrasound. We incorporated this feature with other features in the CADx system and we show that the use of coronal spiculation features (Az = 0.93) significantly improves CADx performance (Az = 0.90, p = 0.02). In addition, a number of morphological features such as height-to-width ratio, compactness, and sphereness are extracted as the shapes of cancers and benign lesions differ. These morphological features strongly rely on the lesion segmentation which is challenging in ultrasound. In this thesis, we used spiral- scanning based dynamic programming to perform the task. However other segmentation methods are also possible such as graph cut, smart opening and level-set. To choose the most appropriate segmentation method in ultrasound, comparison between different segmentation methods based on diagnostic accuracy is needed in future. Moreover, texture analysis is absent in this CADx system. Incorporating features such as co-occurrence based texture features, local binary patterns and Haar-like features may further improve the CADx system.

In *chapter 3*, we conducted an observer study in which we investigated if reader performance can be improved by using a CADx system and we also compared the reader performance with that of the stand-alone CADx system. The CADx system had an area under ROC curve (AUC) of 0.92 for discriminating benign and malignant lesions, where the unaided reader AUC ranged from 0.77 to 0.92. Mean performance of inexperienced readers improved when CADx was used (AUC = 0.85 vs 0.90; p = 0.007), while mean performance of experienced readers did not change with CADx (AUC=0.89). The results shows that our CADx system needs to be improved to be able to aid experienced radiologists. Moreover, the ABUS CADx is a new technique, readers might need more training with CADx to get more confidence in the system and adjust themselves to use the system optimally. In this work, the effect of using CADx on the reading time for each reader is not studied. The time effect (shortening or prolonging the reading time) might be an important factor when incorporating computer systems in the clinical workflow.

For the development of a computer-aided detection (CADe) system, it is necessary to segment the chestwall in ABUS to remove the false positives beyond the chestwall. In *chapter 4*, we present an automatic method to segment the chestwall. We show that the visible part of the chestwall in an automated 3D breast ultrasound image can be accurately modeled by a cylinder. Therefore, we fit the surface of our cylinder model to a set of automatically detected rib-surface points. The detection of the rib-surface points is performed by applying a classifier using features representing local image intensity patterns and presence of rib shadows. The average mean distance of the annotated points to the segmented chest wall was 5.59±3.08 mm. The average result is promising. However, in several cases, our method over-segments thoracic volumes covering a part of breast tissue which means that there is a risk of eliminating cancers from the detection.

In *chapter 5*, we reported a multi-stage detection system for breast cancers in ABUS. Segmentations of the breast, the nipple and the chestwall are performed, providing landmarks for the detection algorithm. Subsequently, voxel features characterizing coronal spiculation patterns, blobness, contrast, and depth are extracted. Using an ensemble of neural-network classifiers, a likelihood map indicating potential abnormalities is computed. Local maxima in the likelihood map are determined and form a set of candidates in each image. These candidates are further processed in a second detection stage, which includes region segmentation, feature extraction and a final classification. On region level, classification was performed to reduce false positives. The computer-aided detection system we developed achieved good detection performance (a view-based sensitivity of 64% at 1 false positives per image).

To improve localization and detection performance of readers, and to facilitate the further use of the CADe system, in *chapter 6* we proposed a method that allows radiologists to locate potential abnormalities in other views (target images) when a lesion has already been pointed out by a reader or CADe in the current view (source image). We combine features describing the lesion location with respect to the nipple, the transducer and the chestwall, with features describing lesion properties such as intensity, spiculation, blobness, contrast and lesion likelihood. By using a grid search strategy, in the feature space, the voxel in the target view which has the minimal distance to the source point was chosen as a target point. Our method achieved an error of 15.64 mm±16.13 mm. The error is small enough to help locate the lesion with minor additional interaction. In the future, we will improve the method by incorporating features extracted at the region level, i.e. the size of the region, acoustic posterior behavior, etc. With those region features, a further improvement is expected, and by using more advanced prediction methods such as supervised learning we could weight the influence of the different features on the result. A limitation of our work is that our method is only applicable to link abnormalities. Full voxel to voxel correspondences between views are not available. A potential solution might be applying a dedicated biomechanical breast model[65] to simulate the compressions of different views from which the correspondences of every voxel across ABUS views can be determined.

To summarize, in this thesis we have developed techniques to automatically segment breast foreground, detect the nipple, segment chestwall, automatically detect, segment and accurately diagnose breast lesions or suspicious regions in automated 3D breast ultrasound images. We demonstrated that using a computer-aided classification system can improve diagnostic performance of inexperienced readers.

Most ultrasound image analysis techniques introduced here rely on the intensity level of voxels in ABUS. To make the techniques robust to ABUS images generated from different systems or manufactures, we performed a step of intensity normalization prior to other image analysis. Currently the normalization is based on simply scaling the intensity level of a breast to a certain range. This method is not ideal, as in ultrasound, the intensities of the same type tissue might differ at different locations. Normalization methods[32] based on separating the image into different frequency bands might be an alternative.

Incorporating computer-aided system into clinical workflow is challenging. First, the computer-aided systems must be automated or require as little intervention from radiologists as possible. Secondly, if the detection system is used as a second reader in screening, the system must be very sensitive to cancers (hardly miss any cancers). About 12 % cancers were missed in the first stage of lesiion candidate detection. Most of the missed lesions are subtle in the coronal planes but still visible in the transversal planes. More robust voxel features characterizing cancers in the transversal planes need to be extracted or even a dedicated detection system for detecting cancers in transversal planes needs to be developed.

Moreover, due to the time limit on radiological reading, the cancer detection system should provide a limited number of CADe marks. In mammography commercial CADe systems currently have about 1 to 3 marks per case[60,62]. Because we are dealing with volumetric data, display of more markers might be acceptable. While an appropriate setting has to be determined experimentally, we believe that more than two marks per volume will not be tolerated in practice. To achieve good performance with limited CADe marks, the number of false positives needs to be further reduced.

A part of our false positives originate from the shadows of ribs. The current chestwall segmentation is not accurate enough to limit our detection within the breast tissue, because our method may over-segment the chest wall which results in a risk that cancers inside the breast tissue are also removed due to an oversegmented chestwall. The failure of those cases are mainly due to false positives generated from rib-surface detection. To obtain a more accurate segmentation, further suppression on those false positives needs to be investigated. Furthermore, in the chestwall segmentation, the fitting procedure only relies on detected rib surface points while the regional information is not considered. In the future, a fitting cost related to likelihoods of voxels being a part of thoracic volume can be adopted to make the method more robust and less sensitive to the outliers generated from rib-surface detection. Additionally, based on our observation, when the scanning depth is sufficiently deep, the rib shadows and the chestwall are hardly visible in the image. For these cases, it is not necessary to perform chestwall segmentation. For future research, criteria or features such as PCA-SIFT[52] which determine the existence of a certain object based on an orientation histogram of pixels inside a region of interest can be studied to determine the necessity to perform the chestwall segmentation.

The false positive reduction also relies on the region classification in the detection system. The classification is cancers and non-lesion regions which includes different types of false positives such as artifacts, rib shadows, fatty tissues, nipple shadows, ducts, etc. A scheme to train a classifier or multiple classifiers to separately deal with several sub-classes might further improve our CADe performance.

In the current system, the computation of co-occurrence matrix texture features relies on the choice of the parameters (offsets). Further experiments can be performed using different parameters to help us optimally use co-occurrence matrix texture features. Additionally, other types of texture features may also be used. For example, local binary pattern (LBP) has been used as an effective descriptor in pattern recognition and texture classification[74]. Sørensen et al.[90] used LBP to automatically classify different texture patterns in lung computed tomography. Incorporation of texture features such as LBP or Haar-like features might be beneficial to our system.

Besides improving detection performance in each individual ABUS images, another way to enhance patient-based sensitivity of the current detection system is to utilize the linkage between detections of the same regions on different ABUS views and perform another stage of classification on each linked region. Samulski and Karssemeijer[87] showed that by linking detected regions in MLO and CC views of the same breast in mammography, the CADe system performance can be significantly improved. In future, similarly, to boost detection performance in ABUS, we could automatically link detected CADe regions using the method introduced in *chapter 6*. In addition, we could link the detected regions in mammography and ABUS to further enhance the region classification performance. The accuracy of the linking algorithm mainly depends on the accuracy of the nipple detection. In the latest ABUS system, nipple position can be extracted from DICOM image header provided by operators.

The introduced techniques in this thesis have a potential to be used in the clinical workflow. However, how to combine those computer systems in a logical and effective way in the clinic remains unanswered. Using computer techniques should on the one hand, improve the reader performance, and on the other hand the use should make radiological reading less time-consuming. Further observer studies on reader performance and reading time with both use of CADe and CADx systems in a sophisticated way are needed to understand how these systems can best be used in practice.

Samenvatting

Ongeveer 40 procent van de vrouwen heeft dens borstweefsel[8]. Vrouwen met extreem dens borstweefsel hebben een risico op het ontwikkelen van borstkanker dat 4 tot 6 maal hoger is dan vrouwen van dezelfde leeftijd zonder dens borstweefsel of met een zeer lage densiteit van het borstweefsel[7]. Voor deze dense borsten is de sensitiviteit van mammografie voor de detectie van kanker echter maar 30-48%[63]. Voor vrouwelijke BRCA 1/2 genmutatiedragers werd de blootstelling aan diagnostische straling vóór de leeftijd van 30 jaar geassocieerd met een verhoogd risico op borstkanker[78]. Door de verschillen in borstdensiteit en het aan- of afwezig zijn van een genmutatie is het kiezen van een gepersonaliseerde screening zeer belangrijk.

Digitale borst tomosynthese mammografie (DBT) is een veelbelovende nieuwe technologie die speciaal werd ontwikkeld voor de beeldvorming van dense borsten. Een 3D-volume van een borst kan worden gereconstrueerd met behulp van meerdere 2D-projectie beelden. Onderzoek heeft aangetoond dat het gebruik van DBT in combinatie met mammografie het aantal heropgeroepen vrouwen in borstkankerscreening kan worden vermindert[38] en de sensitiviteit kan worden verbeterd[3]. Echter, deze technologie is vrij nieuw en meer onderzoek is nodig naar de voordelen van het toepassen van DBM op gepersonaliseerde screening. Daarbij is de dosis die wordt gebruikt bij DBT hoger dan die bij mammografie. Daarom is DBT niet geschikt voor toepassing op BRCA 1/2 genmutatie dragers. Andere studies hebben reeds aangetoond dat het combineren van 2D echografie[6,12] en 3D[37,53] echografie met mammografie de detectie van borstkanker significant verbeterd. Geautomatiseerde 3D borst echografie kan potentieel gebruikt worden als alternatieve screening modaliteit voor vrouwen met dense borsten en wellicht ook voor BRCA 1/2 genmutatiedragers. Toepassing van een gestandaardiseerd imaging protocol maakt geautomatiseerde 3D echografie (ABUS) meer geschikt als een screening modaliteit dan 2D echografie.

Om de gehele borst te scannen worden 3-5 ABUS scans genomen, gericht op verschillende gebieden, terwijl bij mammografie meestal slechts 2 (MLO en CC) afbeeldingen worden genomen. Elke ABUS scan is een volumetrische afbeelding die bestaat uit ongeveer 280 transversale plakken terwijl een mammografische scan slechts een 2D projectie is. ABUS screening vereist daarom meer mankracht voor het interpreteren van de beelden in vergelijking met mammografie. Om de workflow te versnellen en de ABUS screening effectiever te maken, kunnen computersystemen een rol spelen in de toekomst. Een computer-aided detection (CADe) systeem kan worden gebruikt om de lokalisatie van verdachte gebieden te vergemakkelijken en fouten van radiologen te verminderen. Een computer-aided diagnosis (CADx) systeem kan worden gebruikt om kwaadaardige laesies te onderscheiden van goedaardige laesies, waardoor het aantal doorverwezen vrouwen met goedaardige laesies kan worden verminderd.

Hoofdstuk 2 behandelt de ontwikkeling van een CADx systeem voor de karakterisering van borstafwijkingen in ABUS. In ons onderzoek hebben we een systeem ontwikkeld dat spiculatie kenmerken van laesies gebruikt in het gereconstrueerde coronale vlak. Spiculatie is beter zichtbaar in het coronale vlak dan in andere vlakken. Daarnaast weten we dat spiculatie sterk is gecorreleerd met een maligne afwijking en is daarom een zeer waardevol kenmerk. Het spiculatie kenmerk ontbreekt in de meeste huidige CADx systemen voor 2D- echografie en 3D echografie . Wij gebruiken het spiculatie patroon samen met andere kenmerken in ons CADx systeem en laten zien dat het gebruik van spiculatie kenmerken (Az = 0.93) de prestaties van CADx aanzienlijk verbetert (Az = 0,90, p = 0,02). Daarnaast worden een aantal morfologische kenmerken zoals hoogte- breedteverhouding, compactheid en bolvormigheid gebruikt omdat maligne en benigne laesies hierin verschillen. De herkenning van deze morfologische eigenschappen is sterk afhankelijk van de segmentatie van de laesies. In echografie is deze segmentatie zeer gecompliceerd. In dit proefschrift hebben we spiral-scanning based dynamic programming gebruikt om laesies te segmenteren. Echter, andere segmentatiemethoden als graph cut, smart opening en level-set kunnen worden gebruikt. Om de meest geschikte segmentatie methode voor echografie te kiezen is een studie nodig naar de precisie van de verschillende segmentatie methoden. Verder maakt ons CADx systeem geen gebruik van textuur informatie. Het toevoegen van co-occurrence based texture features, local binary patterns and Haar-like features kunnen het CADx systeem mogelijk verder verbeteren.

In *hoofdstuk 3* hebben we onderzocht of het CADx systeem de karakterisatie van borstleasies door radiologen verbetert. Daarnaast hebben we prestatie van het CADx systeem vergeleken met die van de radiologen. De prestatie van zowel het CADx systeem als de radiologen is gemeten middels ROC analyse. De Area under the ROC curve (AUC) voor het onderscheiden benigne en maligne lesies was 0,92 voor CADx en varieerde van 0,77-0,92 voor de radiologen. De gemiddelde prestatie van onervaren lezers verbeterde wanneer CADx werd gebruikt (AUC = 0.85 versus 0.90, p = 0.007), terwijl de gemiddelde prestatie van ervaren lezers niet veranderede met gebruik van CADx (AUC = 0,89) . Het resultaat toont aan dat ons CADx systeem verbeterd moet worden alvorens ervaren radiologen te kunnen helpen. Omdat het ABUS CADx een nieuwe techniek is, kan het zijn dat lezers meer training met CADx nodig hebben, voordat ze het systeem optimaal kunnen benutten. In dit onderzoek is het effect van CADx op de leestijd van de ABUS onderzoeken niet onderzocht. Tijdwinst (verkorten van de leestijd) zou een belangrijke factor zijn bij de integratie van deze computersystemen in de klinische workflow.

Voor de ontwikkeling van een computer-aided detection (CADe) systeem in ABUS, speelt segmentatie van de thoraxwand een belangrijke rol. Correcte segmentatie van de thoraxwand kan het aantal fout positieve bevindingen van CADe omlaag brengen. In *hoofdstuk 4* , presenteren we een automatische methode voor de segmentatie van de thoraxwand. We laten zien dat het zichtbare deel van de thoraxwand in geautomatiseerde 3D borst echografie nauwkeurig kan worden onderscheiden door middel van een cilindervormig model. In deze opzet wordt het oppervlak van ons cilindermodel aan een reeks automatisch gevonden rib-oppervlak punten gebonden. De rib-oppervlak punten worden gevonden op basis van local image intensity patterns en de aanwezigheid van ribschaduwen. De gemiddelde afstand tussen de geannoteerde rib-oppervlak punten en de gesegmenteerde borstwand was 5,59 ± 3,08 mm. Dit resultaat is veelbelovend. Toch komt het in verschillende gevallen voor dat een deel van het borstweefsel binnen de borstwandsegmentatie valt, waardoor de kans ontstaat dat borstafwijkingen niet meer gevonden kunnen worden door het CADe systeem. Daarom moet het systeem in de toekomst nog verbeterd worden.

In *hoofdstuk 5* rapporteerden we een multi-stage detectiesysteem voor borstkanker in ABUS. Segmentaties van de borst, de tepel en de borstwand zijn uitgevoerd en dienden als oriëntatiepunten voor het algoritme. Vervolgens werden voxel karakteristieken zoals coronale spiculatie patronen, blobness, contrast en diepte bepaald. Met behulp van een verzameling van neurale - netwerk classifiers, is een plaatje gemaakt dat de kans op een afwijking aangeeft. Lokale maxima in het waarschijnlijkheidsplaatje werden bepaald en vormden een set van kandidaten in elk beeld. Deze kandidaten werden verder verwerkt in een tweede fase detectie, welke regionale segmentatie, feature extractie en een eindclassificatie omvat. Op een regionaal niveau werd de classificatie uitgevoerd om fout positieven te verminderen. Met dit computer-aided detection systeem behaalden we goede detectie prestaties (een beeld-gebaseerde sensitiviteit van 64% met 1 fout positieve per beeld).

Om lokalisatie en detectie prestaties van radiologen te verbeteren en om het verdere gebruik van de CADe -systeem te vergemakkelijken, hebben we in *hoofdstuk 6* een methode voorgesteld die radiologen de mogelijkheid geeft om potentiele afwijkingen te laten zien in andere weergaven (doelbeeld) wanneer een laesie reeds werd opgemerkt, door een lezer of het CADe systeem, in de huidige weergave (bronbeeld). We combineren aspecten die laesies beschrijven ten opzichte van de tepel, de transducer en borstwand met eigenschappen van de laesie zoals sterkte, spiculation, blobness, contrast en laesie waarschijnlijkheid. Door het gebruik van een (grid search) raster zoekstrategie over alle voxel kenmerken werd de voxel in het doelbeeld, dat de minimale afstand tot de bron punt heeft, als richtpunt genomen. Onze methode behaalde een fout van 15,64 mm ± 16,13 mm. De fout is klein genoeg om te helpen bij het lokaliseren van de laesie met kleine extra interactie. In de toekomst zullen we de methode verbeteren door ook te kijken op regionaal niveau, zoals de omvang van de regio, akoestische posterior gedrag, etc. Met deze regionale kenmerken wordt een verdere verbetering verwacht. En door meer geavanceerde predictiemethodes zoals begeleid leren (supervised learning) konden we de invloed van de verschillende kenmerken bepalen op het resultaat. Een limitatie van ons werk is dat onze methode is alleen van toepassing is op het koppelen van afwijkingen. Volledige voxel naar voxel correspondenties tussen weergaven zijn niet beschikbaar. Een mogelijke oplossing kan gevonden worden door toepassing van een biomechanisch model waarmee realistische vervormingen van de borst tijdens compressie kunnen worden gesimuleerd[65] van diverse standpunten, waaruit het corresponderende voxel in ABUS kan worden bepaald.

Samenvattend, in dit proefschrift hebben we technieken ontwikkeld om automatisch de borst en borstwand te segmenteren en de tepel detecteren. Verder hebben we een systeem ontwikkeld dat automatisch verdachte gebieden in de borst detecteert, segmenteert en diagnostiseert in geautomatiseerde 3D borst ultrasound beelden. We hebben aangetoond dat het gebruik van ons computer-aided classificatiesysteem de diagnostische prestaties van onervaren lezers kan verbeteren.

Publications

Papers in international journals

T. Tan, B. Platel, H. Huisman, C.I. Snchez, R. Mus and N. Karssemeijer. “Computer Aided Lesion Diagnosis in Automated 3D Breast Ultrasound Using Coronal Spiculation”, IEEE Transactions on Medical Imaging 2012;31:1034-1042.

T. Tan, B. Platel, R.M. Mann, H. Huisman and N. Karssemeijer. “Chest Wall Segmentation in Automated 3D Breast Ultrasound Scans”, Medical Image Analysis, Medical Image Analysis 2013;17:12731281.

T. Tan, B. Platel, R. Mus, L. Tabar, R. Mann and N. Karssemeijer. ”Computer-aided Detection of Cancer in Automated 3D Breast Ultrasound”, IEEE Transactions on Medical Imaging, IEEE Transactions on Medical Imaging 2013;32:1698-1706.

T. Tan, Bram Platel, Ritse M. Mann, Thorsten Twellmann, Guido van Schie, Roel Mus, Andr Grivegne and Nico Karssemeijer . ”Evaluation of the Effect of Computer-Aided Classification of Benign and Malignant Lesions on Reader Performance in Automated 3D Breast Ultrasound”, Academic Radiology, Academic Radiology 2013;20(11):1381-1388.

Papers in conference proceedings

T. Tan, H. Huisman, B. Platel, A. Grivignee, R. Mus and N. Karssemeijer. ”Classification of Breast Lesions in Automated 3D Breast Ultrasound”, in: Medical Imaging, volume 7963 of Proceedings of the SPIE, 2011, page 79630X.

T. Tan, B. Platel, T. Twellmann, G. van Schie, R. Mus, A. Grivegnee, L. Tabar and N. Karssemeijer. ”Computer aided interpretation of lesions in automated 3D breast ultrasound”, in: MICCAI Workshop: Breast Image Analysis, 2011.

T. Tan, B. Platel, H. Huisman and N. Karssemeijer. ”Chest wall segmentation in automated 3D breast ultrasound using a cylinder model”, in: MICCAI Workshop: Breast Image Analysis, 2011. PDF

T. Tan, B. Platel, R. Mus and N. Karssemeijer. ”Detection of Breast Cancer in Automated 3D Breast Ultrasound”, in: Medical Imaging, volume 8315 of Proceedings of the SPIE, 2012, pages 831505-1-831505-8.

T. Tan, B. Platel, M. Hicks, R.M. Mann and Nico Karssemeijer. ”Finding Lesion Correspondences in Different Views of Automated 3D Breast Ultrasound”, in: Medical Imaging of Proceedings of the SPIE.

Abstract in conference proceedings

N. Karssemeijer, **T. Tan**, B. Platel, T. Twellmann, L. Tabar, A. Grivignee, R. Mus and H. Huisman. ”A Novel System for Computer-aided Lesion Classification in Automated 3D Breast Ultrasound”, in: Annual Meeting of the Radiological Society of North America, 2011

N. Karssemeijer, **T. Tan**, T. Twellmann, G. van Schie, A. Grivignee, L. Tabar, R. Mann and R. Mus. ”Computer Aided Interpretation of Lesions in Automated 3D Breast Ultrasound”, in: Annual Meeting of the Radiological Society of North America, 2011.

J. van Zelst, R. Mus, **T. Tan**, N. Karssemeijer and R. Mann. ”Feasibility of automated 3D breast ultrasound scanning in screening of women with high risk”, in: European Congress of Radiology, 2012.

Bibliography

[1] Miguel Alem?n-Flores, Patricia Alem?n-Flores, Luis Alvarez-Le?n, Rafael Fuentes-Pav?n, and Jos? M Santana-Montesdeoca. Computer vision techniques for breast tumor ultrasound analysis. *Breast Journal*, 14:483–486, 2008.

[2] Andre Victor Alvarenga, Antonio Fernando C. Infantosi, Wagner Coelho A. Pereira, and Carolina M. Azevedo. Assessing the performance of morphological parameters in distinguishing breast tumors on ultrasound images. *Medical Engineering and Physics*, 32:49 – 56, 2010.

[3] Ingvar Andersson, Debra M. Ikeda, Sophia Zackrisson, Mark Ruschin, Tony Svahn, Pontus Timberg, and Anders Tingberg. Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and birads classification in a population of cancers with subtle mammographic findings. *European Radiology*, 18:2817–2825, 2008.

[4] S. M. Astley and F. J. Gilbert. Computer-aided detection in mammography. *Clinical Radiology*, 59:390–399, 2004.

[5] W.A. Berg. Supplemental screening sonography in dense breasts. *Radiologic Clinics of North America*, 42:845–851, 2004.

[6] W.A. Berg, J.D. Blume, J.B. Cormack, E.B. Mendelson, D. Lehrer, M. Böhm-Vélez, E.D. Pisano, R.A. Jong, W.P. Evans, M.J. Morton, M.C. Mahoney, L.H. Larsen, R.G. Barr, D.M. Farria, H.S. Marques, K. Boparai, and ACRIN 6666 Investigators . Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *Journal of the American Medical Association*, 299:2151–2163, 2008.

[7] Norman F Boyd, Helen Guo, Lisa J Martin, Limei Sun, Jennifer Stone, Eve Fishell, Roberta A Jong, Greg Hislop, Anna Chiarelli, Salomon Minkin, and Martin J Yaffe. Mammographic density and the risk and detection of breast cancer. *New England Journal of Medicine*, 356:227–236, 2007.

[8] Norman F Boyd, Lisa J Martin, Martin Yaffe, and Salomon Minkin. Mammographic density. *Breast Cancer Research*, 11 Suppl 3:S4, 2009.

[9] Rachel F Brem, Janet Baum, Mary Lechner, Stuart Kaplan, Stuart Souders, L. Gill Naul, and Jeff Hoffmeister. Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. *American Journal of Roentgenology*, 181:687–693, 2003.

[10] Mireille Broeders, Sue Moss, Lennarth Nyström, Sisse Njor, Hkan Jonsson, Ellen Paap, Nathalie Massat, Stephen Duffy, Elsebeth Lynge, and Eugenio Paci. The impact of mammographic screening on breast cancer mortality in europe: a review of observational studies. *Journal of Medical Screening*, 19:1425, 2012.

[11] L. Broekhuizen. *Diagnostic strategies for chronic obstructive pulmonary disease*. PhD thesis, Utrecht University, The Netherlands, 2012.

[12] W. Buchberger, A. Niehoff, P. Obrist, P. DeKoekkoek-Doll, and M. Dünser. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Seminars in Ultrasound, CT, and MR*, 21:325–336, 2000.

[13] Heang-Ping Chan, Jun Wei, Berkman Sahiner, Elizabeth A. Rafferty, Tao Wu, Marilyn A. Roubidoux, Richard H. Moore, Daniel B. Kopans, Lubomir M. Hadjiiski, and Mark A. Helvie. Computer-aided detection system for breast masses on digital tomosynthesis mammograms: Preliminary experience. *Radiology*, 237:1075–1080, 2005.

[14] Tony F Chan and Michael K Ng. Galerkin projection methods for solving multiple linear systems.

SIAM Journal on Scientific Computing, 21:836–850, 1999.

[15] Jung Min Chang, Woo Kyung Moon, Nariya Cho, Jeong Seon Park, and Seung Ja Kim. Radiologists' performance in the detection of benign and malignant masses with 3d automated breast ultrasound (abus). *European Journal of Radiology*, 78:99–103, 2011.

[16] Ruey-Feng Chang, Kuang-Che Chang-Chien, Etsuo Takada, Chiun-Sheng Huang, Yi-Hong Chou, Chen-Ming Kuo, and Jeon-Hor Chen. Rapid image stitching and computer-aided detection for multipass automated breast ultrasound. *Medical Physics*, 37:2063–2073, 2010.

[17] Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, and Dar-Ren Chen. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Research and Treatment*, 89:179–185, 2005.

[18] Chung-Ming Chen, Yi-Hong Chou, Ko-Chung Han, Guo-Shian Hung, Chui-Mei Tiu, Hong-Jen Chiou, and See-Ying Chiou. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology*, 226:504–514, 2003.

[19] Dar-Ren Chen, Ruey-Feng Chang, Wei-Ming Chen, and Woo-Kyung Moon. Computer-aided diagnosis for 3-dimensional breast ultrasonography. *Archives of Surgery*, 138:296–302, 2003.

[20] Jie-Zhi Cheng, Yi-Hong Chou, Chiun-Sheng Huang, Yeun-Chung Chang, Chui-Mei Tiu, Kuei-Wu Chen, and Chung-Ming Chen. Computer-aided us diagnosis of breast lesions by using cell-based contour grouping. *Radiology*, 255:746–754, 2010.

[21] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Transactions on Medical Imaging*, 27:629–640, 2008.

[22] D. D. Dorfman, K. S. Berbaum, and C. E. Metz. Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Investigative Radiology*, 27:723–731, 1992.

[23] Karen Drukker, Maryellen L Giger, Karla Horsch, Matthew A Kupinski, Carl J Vyborny, and Ellen B Mendelson. Computerized lesion detection on breast ultrasound. *Medical Physics*, 29:1438–1446, 2002.

[24] Karen Drukker, Maryellen L Giger, and Ellen B Mendelson. Computerized analysis of shadowing on breast ultrasound for improved lesion detection. *Medical Physics*, 30:1833–1842, 2003.

[25] Karen Drukker, Charlene A Sennett, and Maryellen L Giger. Automated method for improving system performance of computer-aided diagnosis in breast ultrasound. *IEEE Transactions on Medical Imaging*, 28:122–128, 2009.

[26] Richard O. Duda and Peter E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15:107–111, 1972.

[27] B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.

[28] K. Emerson. *Principles of Internal Medicine*, pages 582–587. McGraw Hill, New York, 1974.

[29] J. Ferlay, D. M. Parkin, and E. Steliarova-Foucher. Estimates of cancer incidence and mortality in Europe in 2008. *European Journal of Cancer*, 46:765–781, 2010.

[30] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, 127:2893–2917, 2010.

[31] J. Friedman, T. Hastie, and R. Tibshirani. Special invited paper. additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28:337–374, 2000.

[32] L. Gallardo-Estrella, B. van Ginneken, and E. M. van Rikxoort. Normalization of CT scans reconstructed with different kernels to reduce variability in emphysema measurements. In *Medical Imaging*, volume 8670 of *Proceedings of the SPIE*, page 86700E, 2013.

[33] B. S. Garra, B. H. Krasner, S. C. Horii, S. Ascher, S. K. Mun, and R. K. Zeman. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrasonic Imaging*, 15:267–285, 1993.

[34] M. K. Garvin, M. D. Abramoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka. Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-D graph search. *IEEE Transactions on Medical Imaging*, 27:1495–1505, 2008.

[35] M. L. Giger, H. Al-Hallaq, Z. Huo, C. Moran, D. E. Wolverton, C. W. Chan, and W. Zhong. Computerized analysis of lesions in us images of the breast. *Academic Radiology*, 6:665–674, 1999.

[36] Fiona J Gilbert, Susan M Astley, Maureen G C Gillan, Olorunsola F Agbaje, Matthew G Wallis, Jonathan James, Caroline R M Boggis, Stephen W Duffy, and CADET II Group. Single reading with computer-aided detection for screening mammography. *New England Journal of Medicine*, 359:1675–1684, 2008.

[37] Vincenzo Giuliano and Concetta Giuliano. Improved breast cancer detection in asymptomatic women using 3d-automated breast ultrasound in mammographically dense breasts. *Clinical Imaging*, 2012.

[38] Xing Gong, Stephen J. Glick, Bob Liu, Aruna A. Vedula, and Samta Thacker. A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging. *Medical Physics*, 33:1041–1052, 2006.

[39] Lars Kai Hansen and Peter Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:993–1001, 1990.

[40] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:610–621, 1973.

[41] Stephen L Hillis, Kevin S Berbaum, and Charles E Metz. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Academic Radiology*, 15:647–661, 2008.

[42] S. V. Hilton, G. R. Leopold, L. K. Olson, and S. A. Willson. Real-time breast sonography: application in 300 consecutive patients. *American Journal of Roentgenology*, 147:479–486, 1986.

[43] Karla Horsch, Maryellen L. Giger, Luz A. Venta, and Carl J. Vyborny. Computerized diagnosis of breast lesions on ultrasound. *Medical Physics*, 29:157–164, 2002.

[44] Karla Horsch, Maryellen L. Giger, Carl J. Vyborny, and Luz A. Venta. Performance of computer-aided diagnosis in the interpretation of lesions on breast sonography. *Academic Radiology*, 11:272–280, 2004.

[45] H. Huisman and N. Karssemeijer. Chestwall segmentation in 3D breast ultrasound using a deformable volume model. In *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, pages 245–256, 2007.

[46] Yuji Ikedo, Daisuke Fukuoka, Takeshi Hara, Hiroshi Fujita, Etsuo Takada, Tokiko Endo, and Takako Morita. Development of a fully automatic scheme for detection of masses in whole breast ultrasound images. *Medical Physics*, 34:4378–4388, 2007.

[47] V. P. Jackson. The role of us in breast imaging. *Radiology*, 177:305–311, 1990.

[48] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman. Global cancer statistics. *CA A Cancer Journal for Clinicians*, 61:69–90, 2011.

[49] S. S. Kaplan. Clinical utility of bilateral whole-breast us in the evaluation of women with dense breast tissue. *Radiology*, 221:641–649, 2001.

[50] N. Karssemeijer. Local orientation distribution as a function of spatial scale for detection of masses in mammograms. In *Information Processing in Medical Imaging*, volume 1613 of *Lecture Notes in Computer Science*, pages 280–293, 1999.

[51] N. Karssemeijer and G. M. te Brake. Detection of stellate distortions in mammograms. *IEEE Transactions on Medical Imaging*, 15:611–619, 1996.

[52] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, volume 2, pages 506–513, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[53] K.M. Kelly, J. Dean, W.S. Comulada, and S.J. Lee. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European Radiology*, 20:734–742, 2010.

[54] Kwang Gi Kim, Jong Hyo Kim, and Byoung Goo Min. Classification of malignant and benign tumors using boundary characteristics in breast ultrasonograms. *Journal of Digital Imaging*, 15 Suppl 1:224–227, 2002.

[55] T. M. Kolb, J. Lichy, and J. H. Newhouse. Occult cancer in women with dense breasts: detection with screening us–diagnostic yield and tumor characteristics. *Radiology*, 207:191–199, 1998.

[56] E. E. Konofagou. Quo vadis elasticity imaging? *Ultrasonics*, 42:331–336, 2004.

[57] D. B. Kopans, J. E. Meyer, and K. K. Lindfors. Whole-breast us imaging: four-year follow-up. *Radiology*, 157:505–507, 1985.

[58] D. Kotsianos-Hermle, K. M. Hiltawsky, S. Wirth, T. Fischer, K. Friese, and M. Reiser. Analysis of 107 breast lesions with automated 3d ultrasound and comparison with mammography and manual ultrasound. *European Journal of Radiology*, 71:109–115, 2009.

[59] Marla R Lander and László Tabár. Automated 3-d breast ultrasound as a promising adjunctive screening tool for examining dense breast tissue. *Seminars in Roentgenology*, 46:302–308, 2011.

[60] Stephanie Leon, Libby Brateman, Janice Honeyman-Buck, and Julia Marshall. Comparison of two commercial CAD systems for digital mammography. *Journal of Digital Imaging*, 22:421–423, 2009.

[61] B. Liu, H.D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. *Pattern Recognition*, 43:280–298, 2010.

[62] Ansgar Malich, Dorothee R Fischer, and Joachim Böttcher. CAD for mammography: the technique, results, current role and further developments. *European Radiology*, 16:1449–1460, 2006.

[63] M. T. Mandelson, N. Oestreicher, P. L. Porter, D. White, C. A. Finder, S. H. Taplin, and E. White.

Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *Journal of the National Cancer Institute*, 92:1081–1087, 2000.

[64] Lina Arbash Meinel, Alan H Stolpen, Kevin S Berbaum, Laurie L Fajardo, and Joseph M Reinhardt. Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *Journal of Magnetic Resonance Imaging*, 25:89–95, 2007.

[65] Thomy Mertzanidou, John Hipwell, M Jorge Cardoso, Xiying Zhang, Christine Tanner, Sebastien Ourselin, Ulrich Bick, Henkjan Huisman, Nico Karssemeijer, and David Hawkes. MRI to x-ray mammography registration using a volume-preserving affine transformation. *Medical Image Analysis*, 16:966–975, 2012.

[66] Metz and Pan. ”proper” binormal roc curves: Theory and maximum-likelihood estimation. *Journal of Mathematical Psychology*, 43:1–33, 1999.

[67] C. E. Metz, B. A. Herman, and J. H. Shen. Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data. *Statistics in Medicine*, 17:1033–1053, 1998.

[68] Hee Jung Moon, Min Jung Kim, Jin Young Kwak, and Eun-Kyung Kim. Probably benign breast lesions on ultrasonography: a retrospective review of ultrasonographic features and clinical factors affecting the bi-rads categorization. *Acta Radiologica*, 51:375–382, 2010.

[69] W. Moon, Y. Shen, M. Bae, C. Huang, J. Chen, and R. Chang. Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images. *IEEE Transactions on Medical Imaging*, 2012.

[70] Woo Kyung Moon, Chung-Ming Lo, Jung Min Chang, Chiun-Sheng Huang, Jeon-Hor Chen, and Ruey-Feng Chang. Computer-aided classification of breast masses using speckle features of automated breast ultrasound images. *Medical Physics*, 39:6465–6473, 2012.

[71] Woo Kyung Moon, Yi-Wei Shen, Chiun-Sheng Huang, Li-Ren Chiang, and Ruey-Feng Chang. Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images. *Ultrasound in Medicine and Biology*, 37:539–548, 2011.

[72] Marilyn J Morton, Dana H Whaley, Kathleen R Brandt, and Kimberly K Amrami. Screening mammograms: interpretation with computer-aided detection–prospective evaluation. *Radiology*, 239:375–383, 2006.

[73] Monika Nothacker, Volker Duda, Markus Hahn, Mathias Warm, Friedrich Degenhardt, Helmut Madjar, Susanne Weinbrenner, and Ute-Susann Albert. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. a systematic review. *BMC Cancer*, 9:335, 2009.

[74] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:971–987, 2002.

[75] Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet*, 2012.

[76] Tracy Onega, Erin J Aiello Bowles, Diana L Miglioretti, Patricia A Carney, Berta M Geller, Bonnie C Yankaskas, Karla Kerlikowske, Edward A Sickles, and Joann G Elmore. Radiologists’ perceptions

of computer aided detection versus double reading for mammography interpretation. *Academic Radiology*, 17:1217–1226, 2010.

[77] N. Otsu. A threshold selection method from gray level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9:62–66, 1979.

[78] Anouk Pijpe, Nadine Andrieu, Douglas F. Easton, Ausrele Kesminiene, Elisabeth Cardis, Catherine Nogus, Marion Gauthier-Villars, Christine Lasset, Jean-Pierre Fricker, Susan Peock, Debra Frost, D Gareth Evans, Rosalind A. Eeles, Joan Paterson, Peggy Manders, Christi J. van Asperen, Margreet G E M. Ausems, Hanne Meijers-Heijboer, Isabelle Thierry-Chef, Michael Hauptmann, David Goldgar, Matti A. Rookus, Flora E. van Leeuwen, EMBRACE , on behalf of G. E. N. E. P. S. O., and H. E. B. O. N. Exposure to diagnostic radiation and risk of breast cancer among carriers of brca1/2 mutations: retrospective cohort study (gene-rad-risk). *British Medical Journal*, 345:e5660, 2012.

[79] Etta D Pisano, Constantine Gatsonis, Edward Hendrick, Martin Yaffe, Janet K Baum, Suddhasatta Acharyya, Emily F Conant, Laurie L Fajardo, Lawrence Bassett, Carl D'Orsi, Roberta Jong, Murray Rebner, and Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353:1773–1783, 2005.

[80] Etta D. Pisano, R Edward Hendrick, Martin J. Yaffe, Janet K. Baum, Suddhasatta Acharyya, Jean B. Cormack, Lucy A. Hanna, Emily F. Conant, Laurie L. Fajardo, Lawrence W. Bassett, Carl J. D'Orsi, Roberta A. Jong, Murray Rebner, Anna N A. Tosteson, Constantine A. Gatsonis, and D. M. I. S. T Investigators Group . Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in dmist. *Radiology*, 246:376–383, 2008.

[81] M.J.D Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7:152–162, 1964.

[82] G. Rahbar, A. C. Sie, G. C. Hansen, J. S. Prince, M. L. Melany, H. E. Reynolds, V. P. Jackson, J. W. Sayre, and L. W. Bassett. Benign versus malignant solid breast masses: Us differentiation. *Radiology*, 213:889–894, 1999.

[83] D. Rotten, J. M. Levaillant, and L. Zerat. Analysis of normal breast tissue and of solid breast masses using three-dimensional ultrasound mammography. *Ultrasound in Obstetrics and Gynecology*, 14:114–124, 1999.

[84] B. Sahiner, H.P. Chan, M.A. Roubidoux, L.M. Hadjiiski, M.A. Helvie, C. Paramagul, J. Bailey, A.V. Nees, and C. Blane. Malignant and benign breast masses on 3d us volumetric images: effect of computer-aided diagnosis on radiologist accuracy. *Radiology*, 242:716–724, 2007.

[85] Berkman Sahiner, Heang-Ping Chan, Marilyn A Roubidoux, Mark A Helvie, Lubomir M Hadjiiski, Aditya Ramachandran, Chintana Paramagul, Gerald L LeCarpentier, Alexis Nees, and Caroline Blane. Computerized characterization of breast masses on three-dimensional ultrasound volumes. *Medical Physics*, 31:744–754, 2004.

[86] M. Samulski, R. Hupse, C. Boetes, R. Mus, G. den Heeten, and N. Karssemeijer. Using Computer Aided Detection in Mammography as a Decision Support. *European Radiology*, 20:2323–2330, 2010.

[87] M. Samulski and N. Karssemeijer. Optimizing Case-based Detection Performance in a Multiview CAD System for Mammography. *IEEE Transactions on Medical Imaging*, 30:1001–1009, 2011.

[88] Y. Sato, C. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, and R. Kikinis. Tissue classi-

fication based on 3D local intensity structures for volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 6:160–180, 2000.

[89] Fritz K W. Schaefer, A. Waldmann, A. Katalinic, C. Wefelnberg, M. Heller, W. Jonat, and I. Schreer. Influence of additional breast ultrasound on cancer detection in a cohort study for quality assurance in breast diagnosis–analysis of 102,577 diagnostic procedures. *European Journal of Radiology*, 20:1085–1092, 2010.

[90] Lauge Sörensen, Saher B Shaker, and Marleen de Bruijne. Texture classification in lung CT using local binary patterns. In *Medical Image Computing and Computer-Assisted Intervention*, volume 11 of *Lecture Notes in Computer Science*, pages 934–941, 2008.

[91] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*, 196:123–134, 1995.

[92] L. Tabár, C. J. Fagerberg, A. Gad, L. Baldetorp, L. H. Holmberg, O. Gröntoft, U. Ljungquist, B. Lundström, J. C. Månson, and G. Eklund. Reduction in mortality from breast cancer after mass screening with mammography. randomised trial from the breast cancer screening working group of the swedish national board of health and welfare. *Lancet*, 1:829–832, 1985.

[93] T. Tan, H.J. Huisman, B. Platel, A. Grivignee, R. Mus, and N. Karssemeijer. Classification of breast lesions in automated 3D breast ultrasound. In *Medical Imaging*, volume 7963 of *Proceedings of the SPIE*, page 79630X, 2011.

[94] T. Tan, B. Platel, H. Huisman, C. I. Sánchez, R. Mus, and N. Karssemeijer. Computer aided lesion diagnosis in automated 3D breast ultrasound using coronal spiculation. *IEEE Transactions on Medical Imaging*, 31:1034–1042, 2012.

[95] T. Tan, B. Platel, R. M. Mann, H. Huisman, and N. Karssemeijer. Chest wall segmentation in automated 3D breast ultrasound scans. *Medical Image Analysis*, 17:12731281, 2013.

[96] Tao Tan, Bram Platel, Roel Mus, and Nico Karssemeijer. Detection of breast cancer in automated 3D breast ultrasound. In *Medical Imaging*, volume 8315 of *Proceedings of the SPIE*, pages 831505–1–831505–8, 2012.

[97] C. Tanner, N. Karssemeijer, and G. Szekely. Deformation models for registering MR and 3D ultrasound breast images. In *Proc. IEEE Int Biomedical Imaging: From Nano to Macro Symp*, pages 582–585, 2011.

[98] Paul Taylor and Henry W W Potts. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *European Journal of Cancer*, 44:798–807, 2008.

[99] S. Timp and N. Karssemeijer. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. *Medical Physics*, 31:958–971, 2004.

[100] S. Timp, N. Karssemeijer, and J. Hendriks. Analysis of changes in masses using contrast and size measures. In *IWDM '02: Proceedings of the 6th international workshop on Digital Mammography*, pages 240–242. Springer-Verlag, 2002.

[101] Sheila Timp, Celia Varela, and Nico Karssemeijer. Computer-aided diagnosis with temporal analysis to improve radiologists' interpretation of mammographic mass lesions. *IEEE Transactions on Information Technology in Biomedicine*, 14:803–808, 2010.

[102] A. Torralba, K.P. Murphy, and W.T. Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:854–869, 2007.

[103] K. A. M. van den Bergh, M.-L. Essink-Bot, E. M. Bunge, E. Th. Scholten, M. Prokop, C. A. van Iersel, R. J. van Klaveren, and H. J. de Koning. Impact of computed tomography screening for lung cancer on participants in a randomized controlled trial (NELSON trial). *Cancer*, 113:396–404, 2008.

[104] Jiahui Wang, Roger Engelmann, and Qiang Li. Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique. *Medical Physics*, 34:4678–4689, 2007.

[105] Ying Wang, Shuangquan Jiang, Hong Wang, Yan Hui Guo, Bo Liu, Yan Hou, Hengda Cheng, and Jiawei Tian. CAD algorithms for solid breast masses discrimination: evaluation of the accuracy and interobserver variability. *Ultrasound in Medicine and Biology*, 36:1273–1281, 2010.

[106] D. O. Watermann, M. Földi, A. Hanjalic-Beck, A. Hasenburg, A. Lüghausen, H. Prömpeler, G. Gitsch, and E. Stickeler. Three-dimensional ultrasound for the assessment of breast lesions. *Ultrasound in Obstetrics and Gynecology*, 25:592–598, 2005.

[107] E. Wenkel, M. Heckmann, M. Heinrich, S. A. Schwab, M. Uder, R. Schulz-Wendtland, W. A. Bautz, and R. Janka. Automated breast ultrasound: lesion detection and bi-rads classification–a pilot study. *Fortschr Röntgenstr*, 180:804–808, 2008.

[108] Daren Yu, Qinghua hu, and Wen Bao. Combining multiple neural networks for classification based on rough set redcution. In *IEEE 2003 International Conference on Neural Networks & Signal Processing*, 2003.

Acknowledgements

When I was about to finish my master study in Eindhoven in summer 2007, I was already determined to do a PhD study in medical image analysis field. After searching, I found an outdated advertisement of a PhD position from Dr. Henkjan Huisman posting in the research website from the radiology department. I tried my luck to contact Henkjan, and he kindly invite me to Nijmegen, a completely new city for me. During the interview, I met Henkjan and Prof. Nico Karssemeijer, my supervisor and promoter. Initially I was told that I am to going to save men's life: prostate image analysis, but on my first official day here, Nico gave me a supervise. I was going to look at women' breasts. My PhD about breast image analysis began.

Nico, I am lucky to be one of your students and thank you for your excellent supervision. You insightful ideas, creativity and endless knowledge on science inspired me. You are very kind and patient. Even you are a professor, your knowledge on practical programming and scripting also benefit me a lot. Not only from the supervision from you but also the conference trip together with you, I learned a lot of research knowledge and general academic knowledge from you. Moreover, I love to hear your anecdotes and your interesting academic and industry-related career which help me to think about my future career after my PhD. Thank you Nico for your guidance in every areas.

When I was in Eindhoven doing my master, I always heard from my friends that how nice Dr. ir. Bram Platel was when he supervised his students. Bram, after I started my PhD, luckily I met you in Nijmegen again. As my daily supervisor and co-promoter, you are always patient when I explain my ideas about the projects and also you gave me many advices and supervision on the projects. Also thank you for spending lots of time on correcting my paper-writing. You also even helped me with plotting fancy pictures for my papers that definitely increase chances of the acceptance. I still remember in one celebration party for the coming of Sinterklaas, you wrote a wonderful poem in a hilarious way about my study and my research project. Thanks a lot!

Dr. Ritse. Mann, you are always easy to approach to and you spend lots of time on answering me questions. Many thanks for being my supervisor and co-promotor. I would like to thank you for your endless knowledge on the breast imaging. You supervision helped me to develop something in a direction which clinically make sense. Moreover, you are an excellent writer, I learned a lot from your corrections on my papers.

Henkjan, thank you for offering me an opportunity to come to Nijmegen for my PhD interview. You helped me a lot with my initial stage of my PhD. Special thanks to Prof. Bram van Ginneken. Bram, your concept of challenges in medical image analysis and ideals about this group inspired me very well. In the first year of my PhD, Dr.

Michiel Kallenberg gave lots of help on software and programming and we had nice visits to each other's places and cooked for each other. Thank you for your help. When I started to work in this group, I was the only foreigner student. Later on I met Albert who is a pro in making mojito and also an excellent researcher. Albert, after you come here for your master project, you make me not the only foreigner student, and we had lots of fun together. Later on Clarisa (although she is a faculty), Pragnya joined the group, the number of foreigners grows. Thank you guys for inviting me to your parties and enjoyed the life in this group and in Holland.

Specially thanks to Prof. David Hawkes for hosting me for my research visit in the CMIC group, UCL, London. A word of thanks should go to Dr. Roel Mus, Dr. Mathijn de Jong, Dr. Matthieu Rutten, Dr. László Tabár and Dr. André Grivegnée, thank you for your generous help with your precious data and the help with our observer studies and research. Many thanks to Dr. Wei Zhang for helping me with the projects and providing valuable guidance on CAD development.

Many thanks to work partners. Clarisa helped me with pattern recognition topic which benefit me a lot in my paper. Bjoern Eiben from CMIC group, UCL, his experience with registration and bio-mechanical modeling make our collaborated work on model-based ABUS registration happen. Also special thanks to other breast ultrasound team members: Mandana, and Jan. I met Dr. Jan van Zelst when I was preparing our CADe observer study for HAMAM project. I was really in lots of stress. But Jan was really helpful with data selection and testing the workstation. Later on, we worked together on a number of projects and prepared new observer studies. Jan, you are a funny guy and you don't mind I gave you a nick name, flappie. You bring laugh to this group.

Dear DIAG colleagues, thank you for the wonderful time together. I enjoyed my time with all you at work and at DIAG weekend, DIAG football matches and many conference trips. I enjoyed the work and fun. Many thanks to Leonie. Thank you so much for helping me many many times with my VISA applications, preparing letters and lots of paper work. Special thanks to Mark, Jurre and Steven for helping me with my dutch language.

I would like to thank my family all, for your everlasting trust, encouragement, and support towards me.

Nian, thank you for being with me, without you I would never have finished this thesis. Thanks for your patience and believe and thank you for the good life and happiness together with me.

Special thanks to my buddies in Holland and China: Tao, Shiqi, Anmin, Xi, Yi, Minglu, Luc, Ziye, Ling, Nan, Shengsheng, Zhaohui, and many many more. I love to have fun with you guys and enjoy the life.

Curriculum Vitae

Curriculum Vitae

Tao Tan was born in Wuhu, People's Republic of China, on 7 March 1984. In 2002, Tao Tan started studying Biomedical Engineering (Electronic Engineering direction) for his bachelor study at the Zhejiang University, in Hangzhou, China. From 2007 to 2009, he studied Biomedical Engineering at the Eindhoven University of Technology (TU/e) for his master study. January 2010 he started as a Phd Student in the diagnostic image analysis group, the department of radiology, radboud university nijmegen medical centre. The results of the work he carried out in the Diagnostic Image Analysis Group of that department are described in this thesis.